

BE



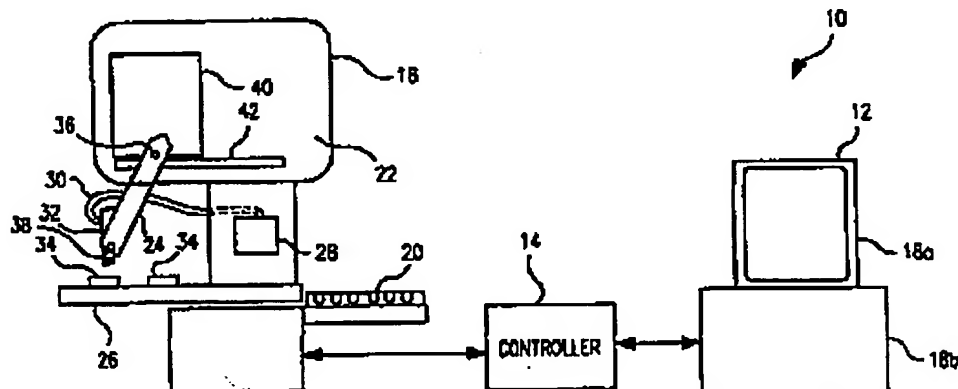
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|---|--|---|---|
| (51) International Patent Classification: C12Q 1/68 | | A1 | (11) International Publication Number: WO 99/31278 |
| | | | (43) International Publication Date: 24 June 1999 (24.06.99) |
| (21) International Application Number: PCT/US98/267:8 | | | |
| (22) International Filing Date: 15 December 1996 (15.12.96) | | | |
| (30) Priority Data: 08/990,851 15 December 1997 (15.12.97) US | | | |
| (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 08/990,851 (CIP) Filed on 15 December 1997 (15.12.97) | | | |
| (71) Applicants (for all designated States except US): SEQUENOM, INC. (US/US); 11555 Sorrento Valley Road, San Diego, CA 92121 (US); KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (KR/KR); 373-1, Yusong-dong, Yuseong-gu, Taejeon 305-701 (KR). | | | |
| (71)(72) Applicant and Inventor: O'DONNELL, Maryanne, J. (US/US); Apartment 9EF, 65 India Row, Boston, MA 02110 (US). | | | |
| (72) Inventors; and (75) Inventors/Applicants (for US only): KANG, Chungwon (KR/KR); Expo Apartment 305 401, 464-1, Juseon- | | | |
| | | min-dong, Yuseong-gu, Taejeon 305-290 (KR); KWON, Young Soo (KR/KR); 25-2, Mangwol-dong, Puk-gu, Kwangju 500-500 (KR); KIM, Young, Tae (KR/KR); Simpson Apartment 2-204, 162, Songpa 2-dong, Songpa-gu, Seoul 138-172 (KR); KOSLER, Herbert (DE/US); 5636-C Via Mallorca Drive, La Jolla, CA 92037 (US); LITTLE, David, P. (US/US); 393 Glendale Lake Road, Patton, PA 18668 (US); XIANG, Huobing (CN/US); 8655 Andromeda Road, San Diego, CA 92121 (US); LUGG, David, M. (GB/GB); 32 Deanehead Road, Eyemouth, Berwickshire TD14 5SA (GB); CANTOR, Charles (US/US); 11 Bay State Road #5, Boston, MA 02215 (US). | |
| | | (74) Agent: SEIDMAN, Stephanie, L.; Heller Eisman White & McAffiffe, Suite 700, 4250 Executive Square, La Jolla, CA 92037 (US). | |
| | | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SH, SG, SI, SK, SL, TJ, TM, TR, TT, CA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GW, ML, MK, NE, SN, TD, TG). | |
| | | Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments. | |

(54) Title: **MASS SPECTROMETRIC METHODS FOR SEQUENCING NUCLEIC ACIDS**



(57) Abstract

A mass spectrometric method for sequencing nucleic acids using RNA polymerases, including DNA-dependent and RNA-dependent RNA polymerases, is provided. The methods include use of the RNA polymerase to generate a set of nested RNA transcripts, which are analyzed mass spectrometry. A method of identifying transcriptional terminator sequences or attenuator sequences is also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|---------------------------------------|----|---|----|--------------------------|
| AL | Albania | KS | Kuwait | LR | Laos | SI | Slovenia |
| AM | Armenia | LI | Liechtenstein | LT | Lithuania | SK | Slovakia |
| AT | Austria | LU | Luxembourg | LC | Liberia | SN | Senegal |
| AU | Australia | GA | Ghana | LV | Latvia | SE | Sweden |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BD | Bangladesh | GT | Guatemala | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MY | Malaysia | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NR | Nauru | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Cote d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | KR | Republic of Korea | PL | Poland | | |
| CN | China | KZ | Kazakhstan | PT | Portugal | | |
| CU | Cuba | LC | Saint Lucia | RO | Romania | | |
| CZ | Czech Republic | LJ | Liechtenstein | RU | Russian Federation | | |
| DE | Germany | DK | Denmark | SD | Sudan | | |
| DK | Denmark | LI | Liechtenstein | SE | Sweden | | |
| EE | Estonia | | | SG | Singapore | | |

-1-

MASS SPECTROMETRIC METHODS FOR SEQUENCING NUCLEIC ACIDS RELATED APPLICATIONS

Benefit of priority is claimed to U.S. application Serial
No.08/990,851, filed December 15, 1997, to Changwon Kang, Young-

5 Soo Kwon, Young Tae Kim, Hubert Köster, Daniel Little, Maryanne
O'Donnell, Guobing Xiang, Charles Cantor and David Lough, entitled
"MASS SPECTROMETRIC METHODS FOR SEQUENCING NUCLEIC
ACIDS". Where permitted, the subject matter of this application is
incorporated herein by reference in its entirety.

10 The subject matter of this application is related to the following
copending applications, international applications and patents: U.S.
Patent Nos. 5,605,798, 5,622,824 and 5,547,835, 5,691,141,
International PCT application No. WO 98/20019, WO 98/20166 and
WO 97/42348. Where permitted, the subject matter of each of these
15 applications and patents is incorporated herein in its entirety.

BACKGROUND OF THE INVENTION

Nucleic acid sequencing methods are important and powerful tools
in the molecular biologist's repository of techniques for assessing and
understanding gene expression and regulation. Methods for sequencing
20 DNA molecules include chemical degradation sequencing (Maxam *et al.*
(1977) *Proc. Natl. Acad. Sci. USA* 74: 560, see, also, Ambrose *et al.*
(1987) *Methods Enz.* 152: 522) and chain termination sequencing
(Sanger *et al.* (1977) *Proc. Natl. Acad. Sci. USA* 74: 5463). The
Maxam-Gilbert sequencing method is a degradative method that relies on
25 specific cleavage of DNA fragments. A fragment of DNA, which is
radiolabeled at one end, is partially cleaved in five separate chemical
reactions. Each reaction is specific for a type of base or a base so that
five populations of labeled fragments of differing lengths are generated.

-2-

The populations of fragments are resolved by polyacrylamide gel electrophoresis (PAGE).

The chain termination or Sanger method relies on a DNA synthesis. Single-stranded DNA is used as a template and labeled primers are used
5 to initiate formation of complementary strands. The synthesis reaction is run in the presence of the four deoxynucleotides (dNTPs), dATP, dTTP, dCTP and dGTP and one of the four dideoxynucleotides (ddNTPs), ddATP, ddTTP, ddCTP, and ddGTP. The complementary strands are prematurely terminated along the chain by addition of dideoxynucleotides
10 to the growing chains. By starting with four reaction mixtures, which each contain one of the four ddNTPs, sets of strands terminating in A, C, G and T are generated. Each reaction mixture is electrophoresed in one lane of a polyacrylamide gel that resolves fragments that differ by a single base in length.

15 These methods are primarily designed for sequencing single-stranded DNA, which is produced by denaturing the DNA and separating the single strands, or by other methods, such as cloning into single-stranded phage vectors. These sequencing methods rely on reactions that produce an array of fragments that differ in length by a single base
20 and terminate in an identifiable base. The fragments are resolved by size using PAGE and are detected using a label, such as a radioisotope. Because the resolution of bands on an electrophoretic gel decreases exponentially as the length of the DNA fragments increase, these methods only permit DNA fragments of up to about 300 to 400
25 nucleotides to be sequenced.

The target DNA can also be amplified using the polymerase chain reaction (PCR) (see, e.g., Mullis *et al.* (1986) Cold Spring Harbor Symp. Quant. Biol. 51: 263; U.S. Patent No. 4,683,202 to Mullis *et al.*), which results in an amplified concentration of the duplex target DNA. A

-3-

number of methods have been used to generate single-stranded templates directly from PCR for subsequent sequencing. For example, radio-labeled primer that is specific for only one strand may be used. Alternatively, PCR may be run under conditions such that one primer is at
5 limited concentration. Once the primer that is at the limited concentration is exhausted, the second strand is amplified at a linear rate through succeeding cycles (Gyllenstein et al. (1988) Proc. Natl. Acad. Sci. 85: 7652; Mihilovic et al. (1989) BioTechniques 7:14).

There are numerous sequencing strategies in use. The selected
10 strategy depends upon the purpose for which the DNA is sequenced and the amount of information available about the DNA prior to sequencing. For example, if the target DNA is sequenced in order to confirm that a particular mutation has been introduced into the DNA, it may only be necessary to sequence a small region of DNA. If the DNA fragment is an
15 unknown gene or portion of a gene for which a sequence must be accurately determined, then it may be necessary to sequence the entire fragment. Because the sizes of fragments that can be sequenced are limited to about 400 bases, DNA fragments longer than this size must be cleaved. Cleavage may be random, by subcloning segments of the target
20 DNA. The subcloned fragments, which include overlapping fragments, are then sequenced, and ordered using a computer program (see, e.g., Staden (1986) Nucl. Acids Res. 14:217). Alternatively, the DNA may be systematically subcloned by generating and sequencing overlapping or nested mutants or by other ordered approaches.

25 The Sanger chain termination method and other sequencing methods rely on the use of single-stranded template by cloning the target DNA into single-stranded phage vectors. The use of plasmids as vectors for the target DNA, however, is preferred over the use of phage DNA for reasons, which include the variety of available plasmids, the ease with

-4-

which plasmids are manipulated, and the greater stability of inserted DNA in plasmids compared to phage vectors. Consequently, methods for sequencing in which the target DNA is cloned into plasmid DNA, rather than into a single-stranded phage DNA (see, e.g., Wallance *et al.* (1981)

- 5 Gene 16: 21-26; Guo *et al.* (1982) Nucl. Acids Res. 10: 2065-2084; Vieira *et al.* (1982) Gene 19: 259-268) have been developed.

- These methods are designed to only sequence one strand of the target DNA at a time (see, e.g., Chen *et al.* (1985) DNA 4: 165-170; Hattori *et al.* (1986) Anal. Biochem. 152: 232-238; Mierendorf *et al.* 10 (1987) Methods Enzymol. 152: 556; Mehra *et al.* (1986) Proc. Natl. Acad. Sci. USA 83: 7013-7017). Use of double-stranded DNA, however, avoids the subcloning or isolation of single-stranded DNA fragments, which are used for the dideoxy chain terminator sequencing reactions. The use of double-stranded DNA, however, had been limited 15 because of the poor template quality of denatured duplex DNA. As a result, these methods had not provided as accurate sequence data as provided by methods in which the DNA is cloned into a single-stranded vector.

- Recently, the problems associated with template quality have been 20 solved by the development of methods that use plasmids that include sites, adjacent to the complementary strands of the inserted DNA, to which strand-specific primers may be efficiently hybridized. By virtue of these methods each of the single strands of double-stranded DNA can be sequenced directly from plasmid DNA without prior subcloning into 25 phage vectors (see, e.g., Chen *et al.* (1985) DNA 4: 165-170 and Chi *et al.* (1988) Nucl. Acids Res. 16: 10382). A strand specific synthetic primer is annealed to covalently closed circular DNA, which has been denatured by heat or alkali, before proceeding with dideoxy sequencing reactions. Alternatively, the primer can be annealed to open circle

-5-

double-stranded plasmid DNA, which has been denatured by alkali, as a template (see, Hattori *et al.* ((1986) Anal. Biochem. **152**: 232-238). The double-stranded DNA is denatured with alkali or heat prior to sequencing using the Sanger method, which is performed at 37° C or higher. The use of different "forward" and "reverse" primers, which are each complementary to the lac Z sequences adjacent to the EcoRI site in λgt11, for separately sequencing each strand of DNA that has been cloned into λgt11 has also been described (see, Mehra *et al.* (1986) Proc. Natl. Acad. Sci. USA **83**: 7013-7017).

10 Plasmids with oppositely oriented promoter regions, which are used in methods which involve transcription, are also used as vehicles for target DNA which is to be sequenced. Each promoter region serves as a distinct specific priming site for sequencing the inserted DNA. Such plasmids are commercially available. For example, the twin promoter 15 plasmid pGEM™ contains the bacteriophage SP6 and T7 RNA polymerase promoters in opposite orientations (Mierendorf *et al.* (1987) Meth. Enzymol. **152**: 556-562) which can generate transcripts for sequencing.

Most methods for sequencing nucleic acids require the use of polyacrylamide gel electrophoresis (*i.e.*, PAGE) that can result in 20 sequencing artifacts or require detectable labels, such as radiolabels, enzymes, or fluorescent or chemiluminescent moieties. Methods for sequencing of DNA using mass spectrometric detection are also in use (see, U.S. Patent No. 5,547,835). These methods primarily rely on sequencing of DNA.

25 Using DNA sequencing methodologies, the entire sequence of the human genome ultimately will be determined. The knowledge of the complete sequence of the human genome DNA will help to understand, to diagnose, to prevent and to treat human diseases. To be able to tackle successfully the determination of the approximately 3 billion base

-6-

pairs of the human genome in a reasonable time frame and in an economical way, rapid, reliable, sensitive and inexpensive methods are required.

Therefore it is an object herein to provide additional methods for sequencing. In particular, it is an object herein to provide mass spectrometric methods of sequencing nucleic acids using RNA polymerase. It is a further object herein to provide methods of sequencing nucleic acids in an array format using RNA polymerase in which nucleic acid probes are immobilized to supports at high densities to facilitate mass spectrometric detection. It is also an object herein to provide methods for identifying transcriptional terminator sequences using mass spectrometric methods.

SUMMARY OF THE INVENTION

Methods for sequencing nucleic acids, particularly DNA, using RNA transcripts are provided. In particular, methods for sequencing nucleic acids using RNA polymerases, including DNA-dependent and RNA-dependent RNA polymerases, are provided. RNA transcripts are analyzed by mass spectrometry.

In practicing the methods, a DNA molecule comprising a promoter and target nucleic acid sequence operatively linked to the promoter is transcribed with an RNA polymerase that recognizes the promoter under conditions that generate a nested set of RNA transcripts that include the target is produced. The molecular weight of the transcripts are determined by mass spectrometry to thereby determine the nucleic acid sequence of the target sequence. The DNA molecule that contains the nucleic acid and target is preferably immobilized.

In certain embodiments, the methods use a double-stranded promoter-containing fragment with a single-stranded region at one end,

-7-

which captures a molecule that contains the target DNA. In these embodiments, the target DNA is preferably single-stranded.

In preferred embodiments, the promoter-containing nucleic acid is covalently coupled via the 3'-end of the noncoding (antisense) strand or
5 5'-end of the coding strand to a solid support. Coupling can be by any method known to those of skill in the art, such as via a thiol linkage. For example, 5'- or 3'-thiolated DNA is prepared and linked at to a aminosilane-treated solid support. The linkage may be direct or via a linker group. The resulting immobilized molecules are preferably arranged
10 in an array format.

A double-stranded nucleic acid molecule encoding a promoter sequence can be obtained from any source, such as bacteria, viruses, bacteriophages, plants and eukaryotic organisms, or assembled by synthetic protocols. It is engineered using so that the resulting fragment
15 contains a single-stranded region of at least a 5 nucleotides at the 3'-end of the coding (sense) strand. This single-stranded region is designed such that it is complementary to a region of the nucleic acid to be sequenced or to a common overlapping sequence, such as a restriction endonuclease site.

20 The target nucleic acid, which includes the region of nucleic acid to be sequenced is single-stranded or double-stranded but includes a single-stranded 3'-end, is hybridized to the complementary sequences of the promoter-containing DNA. Hybridization is performed either before or after immobilization. The hybridization of the two nucleic acid
25 molecules introduces one or more "nicks" in the hybrid at the junction(s) of the adjacent nucleic acid molecules. In certain embodiments, nicks in the coding or non-coding strand, preferably the coding strand, are ligated by the addition of an appropriate nucleic acid ligase prior to initiating transcription.

-8-

- The methods preferably employ any suitable strategy for generating nested RNA transcripts. The preferred methods rely on a modified Sanger sequencing strategy in which RNA polymerase is used to generate a set of nested RNA transcripts obtained by base-specific chain termination. These are analyzed by mass spectrometry. A modified Gilbert-Maxam strategy is also contemplated. In this method, the transcripts are cleaved with base-specific RNases to produce base-specifically terminated fragments and the molecular weights are then assessed by mass spectrometry.
- Transcription is initiated from the promoter by the addition of the appropriate RNA polymerase in the presence of ribonucleoside triphosphates and a selected base-specific chain terminating 3'-deoxyribonucleoside triphosphate. In preferred embodiments, the transcription mixture also contains modified ribonucleoside triphosphates and ribonucleotide analogs ribonucleotide triphosphate analogs that reduce secondary structure of the resulting RNA transcripts and/or increase the fidelity of termination and turnover of the RNA polymerase enzyme thereby increasing the amount of RNA transcript available for analysis.. These analogs include, but are not limited to: inosine 5'-triphosphate (ITP), which reduces secondary structure, 4-thio uridine 5'triphosphate (UTP) and 5-bromo UTP or 5'-iodo CTP, which increase the fidelity of termination and turnover of the RNA polymerase enzyme thereby increasing the amount of RNA transcript available for analysis. The resulting RNA transcripts are analyzed by mass spectrometry.
- In other embodiments, the sample further contains a matrix material for mass spectrometry for analysis by matrix-assisted laser desorption/ionization mass spectrometry (MALDI) and preferably further uses time-of-flight (TOF) analysis. The sequence of the nucleic acid is obtained by aligning the observed mass of the chain-terminated RNA

-9-

transcripts obtained from sequencing reactions containing each of the four chain-terminating bases.

- The methods of sequencing provided herein may be used for diagnostic applications. Diagnostic applications, include methods for
- 5 determination of the presence of genetic alterations in a known target nucleic acid. For example, a region of the target nucleic acid is amplified and the nucleic acid strand corresponding to the noncoding strand is isolated. The nucleic acid probe containing the promoter may be isolated from a natural source or assembled synthetically by hybridizing two
- 10 complementary oligonucleotides to form a promoter sequence. A single-stranded region of at least 5 nucleotides that is complementary to a region of the nucleic acid to be sequenced or to a common sequence is introduced by recombinant means at the 3'-end of the coding strand. In preferred embodiments, the promoter-containing nucleic acid is
- 15 covalently coupled via the 3'-end of the noncoding strand or 5'-end of the coding strand to a solid support and, more preferably, is a 5'- or 3'-thiolated DNA linked at high densities to a aminosilane-treated solid support. The linkage may be in the absence or presence of a linker group and is preferably arranged in an array format.
- 20 A single-stranded 3' overhang of the nucleic acid to be sequenced, in single-stranded or double-stranded form, is hybridized to the complementary sequences of the noncoding strand and, in some embodiments, the nick(s) between one or more nucleic acid strands is/are ligated prior to transcription. Transcription is initiated using the
- 25 appropriate RNA polymerase in the presence of ribonucleoside triphosphates and a selected base-specific chain terminating 3'-deoxyribonucleoside triphosphate. In preferred embodiments, the transcription mixture may also contain inosine 5'-triphosphate and/or one or more modified ribonucleoside triphosphates to facilitate analysis of th

-10-

RNA transcripts. The RNA transcripts are analyzed by mass spectrometry, preferably using MALDI-TOF.

When used in array formats, a panel of promoter-containing nucleic acid probes can be constructed and linked in an array such that
5 the target nucleic acid is permuted along the entire sequence, e.g., the coding sequence of a gene, allowing for the determination of the nucleic acid sequence of the entire gene during a single reaction sequence.

Methods of identifying transcriptional terminator and attenuator sequences are also provided. By modifying transcription conditions to
10 produce full-length transcripts, transcriptional terminator sequences, such as rho-dependent and rho-independent terminators, may be identified using mass spectrometric methods. In practicing the methods, a single-stranded region of the 3'-end of the target nucleic acid to be sequenced is hybridized to a complementary sequence at the 3'-end of the coding
15 strand a promoter-containing nucleic acid probe. In preferred embodiments, the promoter-containing nucleic acid is covalently coupled via the 5'-end of the noncoding strand or 3'-end of the coding strand to a solid support and, more preferably, is a 5'- or 3'-thiolated DNA linked at high densities to a aminosilane-treated solid support. The linkage may be
20 in the absence or presence of a linker group and is preferably arranged in an array format.

Transcription is initiated in the absence or presence of modified RNA triphosphate analogs that increase the efficiency of RNA polymerase termination at such terminator sequences, such as 4-thio UTP, 5-bromo
25 UTP or 5'-iodo CTP. In certain embodiments, nicks in one or more strand may be ligated by the addition of an appropriate nucleic acid ligase prior to initiating transcription (i.e., adding a DNA or RNA ligase). The mass of the terminated RNA transcripts is determined by mass spectrometry. The observed mass is indicative of the location of the terminator-dependent

-11-

arrest of transcription and by comparing the alignment of the sequence immediately preceding the site of transcriptional termination from distinct genomic locations, terminator and attenuator sequences may be identified for different RNA polymerases.

5 DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a system for preparing arrays of a sample material for analysis.

Figure 2 illustrates a pin assembly suitable for use with the system depicted in Fig. 1 for implementing a parallel process of dispensing
10 material to a surface of a substrate.

Figure 3 depicts a bottom portion of the assembly shown in Fig. 2.

Figure 4 depicts an alternative view of the bottom portion of the pin assembly depicted in Fig. 2.

Figures 5A-5D depict a method for preparing an array of sample
15 material.

Figures 6A-6B depict an alternative assembly for dispensing material to the surface of a substrate.

Figure 7 is a schematic showing covalent attachment of oligodeoxynucleotides to a silicon dioxide surface as described herein. In
20 particular, silicon dioxide was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amino groups on the surface. A heterobifunctional crosslinking agent was then reacted with the primary amine to incorporate an iodoacetamido-group. An oligodeoxynucleotide containing a 3'- or 5'-disulfide (shown as the 5') was treated with tris-(2-
25 carboxyethyl) phosphine (TCEP) to reduce the disulfide to a free thiol, which was then coupled to the iodoacetamido-surface.

-12-

Figure 8 is a graph which plots conjugation of oligodeoxynucleotide probes to a silicon surface as a function of TCEP concentration used in the disulfide reduction.

Figure 9 is a matrix assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrum of a silicon wafer with the oligodeoxynucleotide sequence denoted "TCUC" (5'-GAATTCGAGCTCG GTACCCGG-3'; SEQ ID NO: 1) covalently bound essentially as described in Figure 7 and the oligodeoxynucleotide sequence denoted "MJM6" (5'-CCGGGTACCGAGCTCGAATTC-3'; SEQ ID NO: 2) hybridized thereto.

Figure 10 depicts one embodiment of a substrate having wells etched therein that are suitable for receiving material for analysis.

Figure 11 depicts one example of spectra obtained from a linear time of flight mass spectrometer instrument and representative of the material composition of the sample material on the surface of the substrate depicted in Fig. 10.

Figure 12 depicts molecular weights determined for the sample material having spectra identified in Fig. 11.

Figure 13 is a schematic of a 4 x 4 (16-location) DNA array on the surface of a silicon wafer with the thiol-containing oligonucleotide molecules denoted "Oligomer 1", (5'-CTGGATGCGTCGGATCATCTTTT(S)-3'; SEQ ID NO: 5), Oligomer 2 (5'-(S)-CCTCTTGGAAGTGTAGTATT-3'; SEQ ID NO: 6) and "Oligomer 3" (SEQ ID NO: 1; a free thiol derivative "TCUC" oligonucleotide of EXAMPLE 1) covalently bound to 16 locations on the surface of the silicon wafer essentially as described in EXAMPLE 2.

Figure 14 is a schematic of the hybridization of specific oligonucleotides to each of the 16 locations of the DNA hybridization array of Figure 13 with the Oligomer 1 complementary oligonucleotide (5'-GATGATCCGACGCATCAGAATGT-3'; SEQ ID NO: 7) bound to

-13-

Oligomer 1, the Oligomer 2 complementary oligonucleotide (5'-AATACTACACAG-3'; SEQ ID NO: 8) bound to Oligomer 2 and the Oligomer 3 complementary oligonucleotide (5'-CCGGGTACCGAGCTCGAATTC-3'; SEQ ID NO: 2) bound to Oligomer 3.

5 Figure 15 is a representative MALDI-TOF mass spectrum of a 4 x 4 (16-location) DNA array on a silicon wafer shown schematically in Figure 15. The spectrum reveals a single, predominant signal of an experimental mass-to-charge ratio in each location corresponding to the specific hybridized oligonucleotides. The 2+ indicates the position of a
10 doubly charged molecule used as a reference standard during MALDI-TOF MS analysis. The * denotes residual amounts of contaminating oligonucleotide that remain on the surface of the chip following washing procedures. The relative position of the * signal reveals the approximate size of the contaminating oligonucleotide.

15 Figure 16 is a representative MALDI-TOF mass spectrum of an 8 x 8 (64-location) DNA array. The spectrum reveals a single, predominant signal of an experimental mass-to-charge ratio corresponding to the predicted specific hybridized oligonucleotides. The * denotes residual amounts of contaminating oligonucleotide that remain on the surface of
20 the wafer following washing procedures. The relative position of the * signal reveals the approximate size of the contaminating oligonucleotide.

 Figure 17 shows the nucleotide sequence of a DNA molecule (SEQ ID No: 10) assembled by hybridizing a 55-mer oligonucleotide to a complementary 25-mer oligonucleotide (SEQ ID No: 11) and a
25 complementary 30-mer oligonucleotide (SEQ ID No: 12). The resulting double-stranded DNA encodes a SP6 promoter (nt 1-18 of SEQ ID No: 10) and has a single nick in the coding strand of the molecule at nt +7 relative to the start of transcription from the SP6 promoter. The position

-14-

of the nick and the start of transcription initiation from the SP6 promoter are indicated.

DETAILED DESCRIPTION AND PREFERRED EMBODIMENTS

Definitions

5 Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications referred to herein are incorporated by reference herein.

As used herein, the term "nucleic acid" refers to oligonucleotides
10 or polynucleotides such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) as well as analogs of either RNA or DNA, for example, made from nucleotide analogs, any of which are in single or double-stranded form and also PNA. Nucleic acid molecules can be synthetic or can be isolated from a particular biological sample using any number of
15 procedures which are well-known in the art, the particular procedure chosen being appropriate for the particular biological sample.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified nucleotides such as phosphorothioate nucleotides and deazapurine nucleotides. A complete
20 set of chain-elongating nucleotides refers to four different nucleotides that can hybridize to each of the four different bases comprising the DNA template.

As used herein, a nucleic acid promoter-containing probe refers to a nucleic acid fragment that includes a double-stranded region encoding a
25 promoter and a single-stranded region that contains at least 5 nucleotides at the 3'-end of the coding (sense) strand relative to the promoter. The at least 5 nucleotides are selected to be complementary to a single-stranded region at the 3'-end of a nucleic acid containing the target nucleic acid of interest. Reference to the 3'-end refers to the antisense

-15-

strand if the target is contained in a double-stranded molecule. When the target is double-stranded the resulting molecule has nicks in the coding and non-coding strands.

Alternatively, the single-stranded region of the nucleic acid promoter-containing probe includes at least about 5 nucleotides at the 5'-end of the noncoding (antisense) strand (relative to the promoter) that are complementary to at least about five nucleotides at the 5'-end of the coding strand of a target nucleic acid. In this embodiment, the target must be double-stranded. Following hybridization, the resulting molecule has nicks in the coding and non-coding strands.

In all embodiments, nicks can be ligated prior to transcription. As shown herein, however, if the nicks are at a sufficient distance, preferably at least about +7 nucleotides, it is not necessary to ligate.

As used herein, the target nucleic acid is the nucleic acid molecule to be sequenced. It can be part of a larger molecule. The molecule that contains the target nucleic acid can contain or can be modified to contain at least about 5 nucleotides whose sequence is designed to be fully complementary to at least 5 contiguous nucleotides at the 3' end of the overhang on the promoter-containing capture probe (or 5' end depending upon the configuration selected). The target nucleic acid can be double-stranded, with a 3' or 5' single-stranded overhang, or single-stranded. RNA polymerases can and do transcribe single-stranded molecules, particularly, once the ternary complex is formed. The target may also be RNA or PNA (protein nucleic acid formed by conjugating bases to an amino acid backbone; see, e.g., Nielsen et al. (1991) Science 254:1497).

As used herein, nucleic acid synthesis refers to any process by which oligonucleotides or polynucleotides are generated, including, but not limited to processes involving chemical or enzymatic reactions.

-16-

As used herein, a base-specifically terminated ribonucleotides is one that generated during transcription by incorporation a nucleotide that results in transcription termination. Base-specifically terminating ribonucleoside triphosphates, which produce base-specifically terminated
5 ribonucleotides, are known to those of skill in the art. Examples of base-specifically terminating ribonucleoside triphosphates include, but are not limited to: 3'-deoxyribonucleoside triphosphates, such as 3'-dGTP, and others described herein and known to those of skill in the art.

As used herein, complementary when referring to two nucleotide
10 sequences, means that the two sequences of nucleotides are capable of hybridizing, preferably with less than 25%, more preferably with less than 15%, even more preferably with less than 5%, most preferably with no mismatches between opposed nucleotides. Preferably the two molecules will hybridize under conditions of high stringency.

15 As used herein: stringency of hybridization in determining percentage mismatch is as follows:

- 1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C
- 3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C

20 It is understood that equivalent stringencies may be achieved using alternative buffers, salts and temperatures.

As used herein, the term "array" refers to an ordered arrangement of members or positions. The array may contain any number of members or positions and can be in any variety of shapes. In preferred
25 embodiments, the array is two-dimensional and contains $n \times m$ members, wherein m and n are integers that can be the same or different. In particularly preferred embodiments, n and m are each 4 or a multiple thereof.

-17-

The term "cross-linking agent" is art-recognized, and, as used herein, refers to reagents that immobilize a nucleic acid to an insoluble support, preferably through covalent bonds. Thus, appropriate "cross-linking agents" for use herein includes a variety of agents that are

5 capable of reacting with a functional group present on a surface of the insoluble support and with a functional group present in the nucleic acid molecule. Reagents capable of such reactivity include homo- and hetero-bifunctional reagents, many of which are known in the art. Heterobifunctional reagents are preferred.

- 10 As used herein, the term "thiol-reactive functionality," refers to a functionality that reacts rapidly or preferably with a nucleophilic thiol moiety to produce a covalent bond, such as a disulfide or thioether bond. In general, thiol groups are good nucleophiles, and preferred thiol-reactive functionalities are reactive electrophiles. A variety of thiol-reactive
- 15 functionalities are known in the art, and include, for example, haloacetyls (preferably iodoacetyl), diazoketones, epoxy ketones, α , β -unsaturated carbonyls (such as, α , β -enones) and other reactive Michael acceptors (including maleimide), acid halides, benzyl halides, and the like. In certain embodiments, a free thiol group of a disulfide can react with a
- 20 free thiol group, such as by disulfide bond formation, including by disulfide exchange. A "thiol-reactive" cross-linking agent, as used herein, refers to a cross-linking reagent (or surface) which includes, or can be modified to include, at least one thiol-reactive functionality. Reaction of a thiol group can be temporarily prevented by blocking with
- 25 an appropriate protecting group, as is conventional in the art (see e.g.,

-18-

T.W. Greene and P.G.M. Wuts "Protective Groups in Organic Synthesis," 2nd ed. John Wiley & Sons, (1991)).

As used herein, a selectively cleavable linker is a linker that is cleaved under selected conditions, such as a photocleavable linker, a chemically cleavable linker and an enzymatically cleavable linker (i.e., a restriction endonuclease site or a ribonucleotide/RNase digestion). The linker is interposed between the support and immobilized DNA.

As used herein, the terms "protein", "polypeptide" and "peptide" are used interchangeably when referring to a translated nucleic acid, such as a gene product.

As used herein, "sample" refers to a composition containing a material to be detected. In a preferred embodiment, the sample is a "biological sample", which refers to any material obtained from a living source, including but not limited to animal, particularly mammals, including humans, plant, bacteria, fungi, protist and virus. The biological sample can be in any form, including solid materials, such as tissue, cells and biopsied material, and biological fluids, including urine, blood, saliva, amniotic fluid, cerebral spinal fluid and mouth wash (containing buccal cells). Preferably solid materials are mixed with a fluid.

As used herein, "substrate" or "solid support" shall mean an insoluble support onto which a sample is deposited according to the materials as described herein. Examples of appropriate substrates include beads, including but not limited to silica gel, controlled pore glass, magnetic, dextran, agarose and cellulose, capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum, copper and silicon), plastic materials including multiwell plates or membranes (e.g., of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), pins, such as arrays of pins suitable for

-19-

combinatorial synthesis or analysis or beads in pits of flat surfaces such as wafers, particularly silicon wafers, with or without plates.

As used herein, RNA polymerase refers to DNA-dependent RNA polymerases and RNA-dependent RNA polymerases. Any RNA polymerase that recognizes a specific promoter sequence and is capable of initiating transcription and elongating a RNA transcript is contemplated within the scope of the term herein. Exemplary RNA polymerases that may be used in the methods provided herein include, but are not limited to those obtained from: 1) archeobacteria, such as Halobacterium,
10 Methanobacterium, Methanococcus, Sulfolobales and Thermoplasma; 2) eubacteria, such as gram negative bacteria, e.g., Escherichia coli and strains of Salmonella and Shigella, gram positive bacteria, e.g., Bacillus subtilis and Staphylococcus aureus; 3) bacteriophages, such as T7, T3, SP6, SP6 nicked and N4; 4) DNA viruses; 5) RNA viruses, such as
15 influenza virus; 6) plants, such as wheat; and 7) eukaryotic RNA polymerase II isolated from fungi, e.g., Saccharomyces cerevisiae and higher eukaryotic organisms, e.g., mammals. Also included within the scope of the term RNA polymerase as used herein is the RNA phage Q β replicase (see, e.g., International PCT application No. PCT/US87/00880
20 and 5,270,184). Eukaryotic RNA polymerases, such as those isolated from eukaryotic tissues, including plants and animals, that recognize eukaryotic promoters, including viral promoters from viruses that infect eukaryotic cells, are also contemplated herein. The T7, T3, and SP6 RNA polymerases have been cloned and their sequences and methods for
25 their purification as well as for purification of the bacterial RNA polymerases and certain eukaryotic DNA dependent RNA polymerases are known. In addition, numerous RNA polymerase commercially available in large quantities, and can also be readily produced from commercially available vectors. The sequences of the promoters for each

-20-

of the above-noted polymerases and other RNA polymerase are also known to those of skill in the art.

As used herein, a promoter region refers to the portion of DNA of a gene that controls expression of DNA to which it is operatively linked.

- 5 The promoter region includes specific sequences of DNA that are sufficient for RNA polymerase recognition, binding and transcription initiation. This portion of the promoter region is referred to as the promoter. In addition, the promoter region includes sequences that modulate this recognition, binding and transcription initiation activity of
10 the RNA polymerase. These sequences may be cis acting or may be responsive to trans acting factors. Promoters, depending upon the nature of the regulation, may be constitutive or regulated. A constitutive promoter is always turned on. A regulatable promoter requires specific signals to be turned on or off. A developmentally regulated promoter is
15 one that is turned on or off as a function of development.

- The promoter can be of a consensus sequence or variant. When a non-wild-type promoter is used, transcription will occur at a rate sufficient to produce a detectable transcript, and is typically at least about 5-10% of the rate at which transcription would have occurred if a wild-type or
20 native promoter had been used by the RNA polymerase to transcribe the nucleic acid in vitro. Each species of RNA polymerase is specific for a particular promoter sequence. Each of the three bacteriophage RNA polymerases (T7, T3, and SP6) has its own specific DNA promoter sequence. After the RNA polymerase has attached to the promoter
25 sequence of the template and transcription has been initiated, the promoter sequence releases the RNA polymerase. The RNA polymerase then continues transcription along the template until (1) it runs out of RNA starting materials, (2) the transcription reaction is terminated by a particular nucleotide, nucleoside or analog, (3) transcription is

-21-

spontaneously aborted when the RNA polymerase and the template prematurely disassociate or (4) the end of the DNA template is reached.

As used herein, a "promoter-containing nucleic acid" is a nucleic acid that contains a sequence of nucleotides that directs the site-specific binding of an RNA polymerase molecule, which binds and melts the double-stranded DNA to form an open transcription initiation complex that is capable of initiating RNA synthesis in the presence of ribonucleotide triphosphates. Thus, a functional promoter sequence is a sequence of nucleotides to which a DNA dependent RNA polymerase binds, melts in, forms an initiation complex and initiates transcription. The promoter-containing nucleic acid capture probe refers to the double-stranded molecule that includes the promoter region and a 3' overhang of at least about 5 nucleotides to which the target nucleic acid hybridizes, particularly under high stringency conditions. Typically such hybridization requires about 5 contiguous base pairs of identity between the target and promoter-containing capture probe. After initiating transcription, the RNA polymerase (or subunits thereof) is part of ternary complex containing the polymerase, the DNA, and the nascent RNA transcript.

As used herein, a "coding strand" refers to the nucleic acid strand of a promoter-containing nucleic acid that has the same polarity as a corresponding mRNA molecule initiated from that promoter. Thus, it refers to the sense strand. The non-coding strand is the anti-sense strand that is transcribed by RNA polymerase.

As used herein "native nucleoside (or ribonucleoside) triphosphate" or "native nucleotide" or ribonucleotide triphosphate means any one of the naturally occurring normal precursors of RNA, i.e., ATP, GTP, CTP, or UTP. A "nucleoside triphosphate analog" or "nucleotide analog" means any nucleoside triphosphate analogous to a native nucleotide, but that

-22-

contains one or more chemical modifications compared to the native nucleoside or nucleotide.

As used herein, a "matrix material" refers to a material used in mass spectrometry that is a proton donating, UV absorbing material, usually an organic acid, that forms crystalline matrix-nucleic acid structures that are readily ionizable during MALDI. An exemplary matrix material is a solution of 3-hydroxypicolinic acid (3-HPA, 0.7 M in 50% acetonitrile, 10% ammonium citrate).

As used herein, "conditioning" refers to a process whereby prior to mass spectrometric analysis, the nucleic acid to be analysed is modified, for example to decrease the laser energy required for volatilization and/or to minimize fragmentation. Conditioning can be performed on a nucleic acid, preferably after immobilization. An example of conditioning is modification of the phosphodiester backbone of the nucleic acid molecule by, for example, cation exchange), eliminates peak broadening due to a heterogeneity in the cations bound per nucleotide unit. Contacting a nucleic acid molecule with an alkylating agent such as alkyl iodide, iodoacetamide, -iodoethanol, or 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be transformed into a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Conditioning also can be performed by incorporating modified nucleotides into a nucleic acid during synthesis. Such modified nucleotides can be used to reduce sensitivity for depurination (fragmentation during MS), such as with a purine analog such as N5-, N7- or N9-deazapurine nucleotides or ribonucleotide. Modified nucleotides also can include RNA building blocks, oligonucleotide triesters, phosphorothioate functions which are alkylated, or oligonucleotide mimetics such as PNA. In preparing conditioned RNA transcripts with modified nucleotides, the

-23-

RNA polymerase chosen must be able to incorporate the modified nucleotide into the RNA transcript.

Methods for sequencing

Mass spectrometric methods of sequencing nucleic acids are provided. In particular, the sequencing methods use nucleic acid promoter-containing probe that contain a double-stranded region encoding a promoter and a single-stranded region for hybridizing to a complementary region on target nucleic acids. The nucleic acid promoter-containing capture probe is preferably immobilized on a solid support. The single-stranded region is hybridized to a complementary on a target nucleic acid, which contains the nucleic acid to be sequenced and at least 5 nucleotides of single-stranded region identical to the 5 contiguous single-stranded nucleotides on the probe.

Following hybridization of the target to the nucleic acid promoter-containing probes, the nucleic acid sequence is determined by generating a set of nested base-specific chain terminated RNA transcripts initiated from the promoter and transcribed through the target nucleic acid. The resulting nested RNA transcripts are analyzed using mass spectrometry.

While DNA is often the preferred vehicle for sequencing, the methods herein are sequence RNA transcripts. RNA fragments are more stable during matrix-assisted laser desorption/ionization (MALDI) mass spectrometry than DNA fragments. Without being bound by any theory, the enhanced stability may result from the presence of the 2'-hydroxyl group on the sugar moiety of RNA, which helps to reduce depurination during the MALDI process.

-24-

RNA polymerases and promoters

Any RNA polymerase capable of directing the in vitro transcription of an RNA molecule is contemplated for use in the methods described herein. Preferred RNA polymerases are DNA-dependent RNA

- 5 polymerases. Methods for isolating and purifying RNA polymerase molecules are well known to those of skill in the art: Q β replicase (see, e.g., U.S. Patent No. 5,896,249, Re: 35,443 and Eoyang et al. (1971) in Procedures in Nucleic Acid Research, Cantoni and Davies, eds., Volume 2, pp. 829-839, Harper and Rowe, NY); bacteria (e.g., E. coli, see
- 10 Burgess and Jendrisak (1975) Biochemistry 14:4634-4638 and Hager et al. (1990) Biochemistry 29:7890-7894; Leishmania, e.g., see Sadhukhan et al. (1997) Mol. Cell. Biochem. 171:105-114; Bacillus subtilis, Giacomoni (1980) Eur. J. Biochem. 106:579-591); phage (e.g., T7, McDonnell et al. (1977) J. Mol. Biol. 89:719-736 and Studier et al.
- 15 (1989) Virology 39:562-574; see also He et al. (1997) Protein Expr. Purif. 9: 142-151); viruses (e.g., rhabdovirus, see Das et al. (1996) Methods Enzymol. 275:99-122; turnip mosaic virus, see Deidman et al. (1997) J. Viral Meth. 64:184-195; vaccinia, Gershon et al. (1996) Methods Enzymol. 275:35-57); and mammalian (e.g., yeast polII, Koleske
- 20 et al. (1996) Methods Enzymol. 273:176-184; human polII, Maldonado et al. (1996) Methods Enzymol. 274:72-100). In addition, a number of prokaryotic, eukaryotic, bacteriophage and viral RNA polymerases, such as T3, T7 and SP6 are commercially available (e.g., sold by Stratagene, La Jolla, CA; Boehringer Mannheim, Indianapolis, IN; Pharmacia, Uppsala,
- 25 Sweden; and Sigma Chemical Corp, St. Louis, Mo) and/or well known. The promoters are also well known and readily available (e.g., the pGEM series of plasmids are available from Promega, Madison WI; see also, U.S. Patent No. 4,786,072, which describes construction of the pGEM plasmids).

-25-

In practicing the methods, DNA-dependant RNA polymerases and RNA-dependant RNA polymerases may be used. For example, RNA polymerases that may be used in the methods provided herein include, but are not limited to those obtained from: 1) archeabacteria, such as 5 Halobacterium, Methanobacterium, Methanococcus, Sulfolobales and Thermoplasma; 2) eubacteria, such as gram negative bacteria, e.g., Escherichia coli and strains of Salmonella and Shigella, gram positive bacteria, e.g., Bacillus subtilis and Staphylococcus aureus; 3) bacteriophages, such as T7, T3, SP6 and N4; 4) DNA viruses; 5) RNA 10 viruses, such as influenza virus; 6) plants and plant viruses, such as wheat and turnip mosaic virus; and 7) eukaryotic RNA polymerase II isolated from fungi, e.g., Saccharomyces cerevisiae and higher eukaryotic organisms, e.g., mammals (e.g., for a review see In RNA Polymerase and the Regulation of Transcription, Reznikoff et al., eds, Elsevier, NY). Also 15 included for use herein is the Q β replicase from the Q β RNA phage (e.g., see U.S. Patent Nos. 5,670,353, 5,696,249 and Re: 35,443).

The selection of the appropriate RNA polymerase for a nucleic acid template to be sequenced is within the skill of the skilled artisan and varies according to the nucleic acid molecule to be sequenced and the 20 selected promoter region. The selection may be determined empirically following using teachings known to those of skill in the art, including those described herein. Preferred nucleic acid molecules to be sequenced are DNA molecules. RNA and, particularly, PNA (protein nucleic acids) are also amenable to sequencing, upon selection of an 25 appropriate RNA polymerase.

Each nucleic acid promoter-containing probe used in the sequencing methods described herein contains a promoter. The promoters used in the methods herein can be obtained from any source or can be prepared synthetically based upon sequences recognized by

-26-

the selected RNA polymerase. For example, the nucleic acid containing a promoter may be obtained directly from a variety of different organisms, such as bacteria, viruses, phage and eukaryotic organisms, by cloning or may be obtained from commercially available expression vectors (e.g.,

5 T7, T3, SP6 and λ_{PL} and λ_{PR} promoters from Boehringer Mannheim and Pharmacia; bla or lac promoters, RSV-LTR promoter and F9-1 promoter; Stratagene). The selection of the appropriate promoter will depend on the nucleic acid to be sequenced, sequencing conditions, and most importantly, on the RNA polymerase selected for transcription.

10 Promoters can be produced by hybridizing oligonucleotides encoding the coding and non-coding DNA strands of a promoter (see Example 4). The separate strands of nucleic acid are obtained preferably by nucleic acid synthetic techniques. Oligonucleotides used to assemble a promoter are designed such that following hybridization, there exists a
15 single stranded region at the 3'-end of the coding strand to which target DNA can hybridize.

Immobilization of nucleic acid promoter-containing probes

In preferred embodiments, the nucleic acid promoter-containing probe is immobilized, directly or by means of a cross-linking agent, to a
20 solid support provided herein. The probe can be immobilized prior to hybridization or subsequent to hybridization to the target. Preferably and most conveniently hybridization is effected following immobilization to the solid support. The probe is preferably immobilized via the 5' end of the coding strand or the 3' end of the non-coding strand.

25 In other embodiments, a promoter and a downstream target nucleic acid can be isolated as a transcriptional unit from genomic DNA by amplification with primers or hybridization cloning technology. DNA encoding the promoter and target can then be immobilized onto a solid via the terminal nucleotides of the 5' end of the coding strand (sense

-27-

strand) or the 3'-end of the non-coding strand. As discussed further below, a linking group can be incorporated into PCR primers used to isolate the promoter-containing nucleic acid template to achieve directed immobilization to a support. Use of a suitable RNA polymerase that
5 recognizes the promoter, will in the presence of chain terminating nucleotides will result in a set of nested fragments from which the sequence can be determined.

Preferred solid supports are those that can support linkage of nucleic acids thereto at high densities, preferably such that the
10 covalently bound nucleic acids are present on the substrate at a density of at least about 20 fmol/mm², more preferably at least about 75 fmol/mm², still more preferably at least about 85 fmol/mm², yet more preferably at least about 100 fmol/mm², and most preferably at least about 150 fmol/mm². Among the most preferred substrates for use in
15 the particular methods of immobilizing nucleic acids to substrates provided herein is silicon, whereas less preferred substrates include polymeric materials such as polyacrylamide. Substrates for use in methods of producing arrays provided herein include any of a wide variety of insoluble support materials including, but not limited to silica
20 gel, controlled pore glass, cellulose, glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum, silicon and copper), plastic materials (e.g., of polyethylene, polypropylene, polyamide, polyvinylidene-difluoride) and silicon.

In embodiments of the methods in which a cross-linking reagent is
25 not employed, a modified nucleic acid is reacted directly with a appropriately functionalized surface to yield immobilized nucleic acid. Thus, for example, an iodoacetyl-modified surface (or other thiol-reactive surface functionality) can react with a thiol-modified nucleic acid to provide immobilized nucleic acids.

-28-

In embodiments using a cross-linking agent, the cross-linking agent is selected to provide a high density of nucleic acids immobilized on the insoluble support. The cross-linking agent (and other reagents used to functionalize the support surface or the nucleic acid molecule) can be selected to provide any desired spacing of the immobilized nucleic acid molecules from the support surface, and to provide any desired spacing of the immobilized nucleic acids from each other. Thus, steric encumbrance of the nucleic acid molecules can be reduced or eliminated by choice of an appropriate cross-linking agent. In certain embodiments, the cross-linking reagent can be selected to provide multiple reactive functionalities as used in dendrimer synthesis for attachment of multiple nucleic acids to a single cross-linking moiety. Preferably, the cross-linking agent is selected to be highly reactive with the nucleic acid molecule, to provide rapid, complete, and/or selective reaction. In preferred embodiments, the reaction volume of the reagents (e.g., the thiol group reagent and the thiol-reactive functionality) is small.

Modified nucleic acid promoter-containing probes and linkers

Preferred nucleic acid promoter-containing probes for use herein are "thiol-modified nucleic acids," i.e., nucleic acids derivatized to contain at least one reactive thiol moiety. As described in further detail in Example 1, below, nucleic acids containing at least one reactive thiol are preferably made by treating a nucleic acid containing a 3' or 5' disulfide with a reducing agent, which preferably will not compete in subsequent reactions (i.e. will not react with an iodoacetimido functionality). Disulfide-derivatized nucleic acids can be synthesized according to a variety of methods. For example, a nucleic acid can be modified at the 3'- or 5'-terminus by reaction with a disulfide-containing modifying reagent. Alternatively, a thiolated primer can be enzymatically or non-enzymatically attached to the nucleic acid. A 5'-

-29-

phosphoramidate functionality can also provide an attachment point for a thiol or disulfide-containing cytosine or deoxycytosine. Examples of reducing agents appropriate for reduction of a disulfide-modified nucleic acid include: tris-(2-carboxyethyl)phosphine (TCEP) (preferably a concentration in the range of 1-100 mM (most preferably about 10 mM)) is reacted at a pH in the range of 3-6 (most preferably about 4.5), a temperature in the range of 20-45°C (most preferably about 37°C) for a time period in the range of about 1 to about 10 hrs (most preferably for about 5 hrs); dithiothreitol (preferably a concentration in the range of 25 to 100 mM (depending on whether the reactant is isolated) is reacted at a pH in the range of 6-10 (most preferably about 8) and at a temperature in the range of 25-45°C (most preferably about 37°C)) for a time in the range of about 1 to about 10 hrs (most preferably about 5 hrs). TCEP provides an advantage in the low pH at which it is reactive. This low pH effectively protonates thiols, thus suppressing nucleophilic reactions of thiols and resulting in fewer side reactions than with other disulfide reducing agents which are employed at higher pH.

As further described in Example 1, below, a preferred bifunctional cross-linking agent is N-succinimidyl(4-iodoacetyl) aminobenzoate (SIAB). Other crosslinking agents include, but are not limited to, dimaleimide, dithio-bis-nitrobenzoic acid (DTNB), N-succinimidyl-S-acetyl-thioacetate (SATA), N-succinimidyl-3-(2-pyridyldithiol propionate (SPDP), succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) and 6-hydrazinonicotinamide (HYNIC) may also be used in the novel process. For further examples of cross-linking reagents, see, e.g., Wong "Chemistry of Protein Conjugation and Cross-Linking," CRC Press (1991), and Hermanson, "Bioconjugate Techniques" Academic Press (1995).

-30-

In other embodiments, the nucleic acid is immobilized using a photocleavable or photolabile linker moieties that is cleaved during mass spectrometry. Exemplary photolabile cross-linker include, but are not limited to, 3-amino-(2-nitrophenyl)propionic acid (Brown *et al.* (1995)

- 5 Molecular Diversity, pp.4-12 and Rothschild *et al.* (1996) Nucleic Acids Res. 24:361-66) See, also International PCT application No. (WO 98/20019).

- A nucleic acid promoter-containing probe can be directly linked to a solid support via a reversible or irreversible bond between an
- 10 appropriate functionality (L') on the target nucleic acid molecule (T) and an appropriate functionality (L) on the capture molecule. The format is similar to that shown in U.S. Patent No. 5,503,980. A reversible linkage includes linkages that are cleaved under the conditions of mass spectrometry. These include, but are not limited to photocleavable
- 15 bonds, such as a charge transfer complex, and labile bonds formed between relatively stable organic radicals.

- Furthermore, the linkage can be formed with L' being a quaternary ammonium group, in which case, the surface of the solid support, preferably, carries negative charges to repel the negatively charged
- 20 nucleic acid backbone, thereby facilitating the desorption required for analysis by a mass spectrometer. Desorption can occur either by the heat created by the laser pulse and/or, depending on L', by specific absorption of laser energy which is in resonance with the L' chromophore.

- 25 Thus, the L-L' chemistry can be of a type of disulfide bond (chemically cleavable, *e.g.*, by mercaptoethanol or dithioerythrol; a biotin/streptavidin system; a heterobifunctional derivative of a trityl ether group; see, *e.g.*, Köster *et al.* (1990) "A Versatile Acid-Labile Linker for Modification of Synthetic Biomolecules," Tetrahedron Letters 31:7095)

-31-

that can be cleaved under mildly acidic conditions as well as under conditions of mass spectrometry, a levulinyl group cleavable under almost neutral conditions with a hydrazinium/acetate buffer, an arginine-arginine or lysine-lysine bond cleavable by an endopeptidase enzyme like trypsin or a pyrophosphate bond cleavable by a pyrophosphatase, or a ribonucleotide bond in between the oligodeoxynucleotide sequence, which can be cleaved, for example, by a ribonuclease or alkali.

The functionalities, L and L', can also form a charge transfer complex and thereby form the temporary L-L' linkage. Since in many cases the "charge-transfer band" can be determined by UV/vis spectrometry (see, e.g., Organic Charge Transfer Complexes by R. Foster, Academic Press, 1969), the laser energy can be tuned to the corresponding energy of the charge-transfer wavelength and, thus, a specific desorption off the solid support can be initiated. Those skilled in the art will recognize that several combinations can serve this purpose and that the donor functionality can be either on the solid support or coupled to the nucleic acid molecule to be detected or vice versa.

In another approach, a reversible L-L' linkage can be generated by homolytically forming relatively stable radicals. Under the influence of the laser pulse, desorption (as discussed above) as well as ionization will take place at the radical position. Those skilled in the art will recognize that other organic radicals can be selected and that, in relation to the dissociation energies needed to homolytically cleave the bond between them, a corresponding laser wavelength can be selected (see e.g., Reactive Molecules by C. Wentrup, John Wiley & Sons, 1984).

As noted, several embodiments of the method involving of immobilization are contemplated herein. These include, but are not limited to embodiments in which: 1) the primer nucleic acid is immobilized to the solid support and the target nucleic acid is hybridized

-32-

thereto to form a promoter sequence; 2) a double-stranded nucleic acid encoding a promoter (amplified or isolated) is immobilized through linkage to one predetermined strand, and in vitro transcription is initiated in the presence of a predetermined deoxy ribonucleotide, and 3) the target
5 nucleic acid, which can be amplified where there is available sequence information, is hybridized to a double-stranded nucleic acid capture probe with 3' overhang, and the resulting hybrids are immobilized.

In the embodiments where the primer nucleic acid is immobilized on the solid support and the target nucleic acid is hybridized thereto, the
10 inclusion of a cleavable linker permits the primer DNA to be immobilized at the 5'-end so that free 3'-OH is available for "hybridizing" target DNA to the free DNA strand and initiating transcription to

Any linker known to those of skill in the art for immobilizing nucleic acids to solid supports may be used herein to link the nucleic acid
15 to a solid support. The preferred linkers herein are the selectively cleavable linkers, particularly those exemplified herein. Other linkers include, acid cleavable linkers, such as bismaleimideoxy propane and acid-labile trityl linkers. Acid cleavable linkers, photocleavable and heat sensitive linkers may also be used, particularly where it may be
20 necessary to cleave the targeted agent to permit it to be more readily accessible to reaction.

Acid cleavable linkers

Acid cleavable linkers include, but are not limited to, bismaleimideoxy propane; and adipic acid dihydrazide linkers (see, e.g., Fattom
25 et al. (1992) Infection & Immun. 60:584-589) and acid labile transferrin conjugates that contain a sufficient portion of transferrin to permit entry into the intracellular transferrin cycling pathway (see, e.g., Walhöner et al. (1991) J. Biol. Chem. 266:4309-4314).

-33-

Photocleavable Linkers

Photocleavable linkers are linkers that are cleaved upon exposure to light (see, e.g., Goldmacher et al. (1992) Bioconj. Chem. 3:104-107), thereby releasing the targeted agent upon exposure to light.

- 5 Photocleavable linkers that are cleaved upon exposure to light are known (see, e.g., Hazum et al. (1981) in Pept., Proc. Eur. Pept. Symp., 16th, Brunfeldt, K (Ed), pp. 105-110, which describes the use of a nitrobenzyl group as a photocleavable protective group for cysteine; Yen et al. (1989) Makromol. Chem 190:69-82, which describes water soluble
- 10 photocleavable copolymers, including hydroxypropylmethacrylamide copolymer, glycine copolymer, fluorescein copolymer and methylrhodamine copolymer; Goldmacher et al. (1992) Bioconj. Chem. 3:104-107, which describes a cross-linker and reagent that undergoes photolytic degradation upon exposure to near UV light (350 nm); and
- 15 Senter et al. (1985) Photochem. Photobiol 42:231-237, which describes nitrobenzyloxycarbonyl chloride cross linking reagents that produce photocleavable linkages), thereby releasing the targeted agent upon exposure to light. In preferred embodiments, the nucleic acid is immobilized using the photocleavable linker moiety that is cleaved during
- 20 mass spectrometry.

Chemically cleavable linkers

- A variety of chemically cleavable linkers may be used to introduce a cleavable bond between the immobilized nucleic acid and the solid support. Acid-labile linkers are presently preferred chemically cleavable
- 25 linkers for mass spectrometry, especially MALDI-TOF MS, because the acid labile bond is cleaved during conditioning of the nucleic acid upon addition of the 3-HPA matrix solution. The acid labile bond can be introduced as a separate linker group, e.g., the acid labile trityl groups or may be incorporated in a synthetic nucleic acid linker by introducing one

-34-

- or more silyl internucleoside bridges using diisopropylsilyl, thereby forming diisopropylsilyl-linked oligonucleotide analogs. The diisopropylsilyl bridge replaces the phosphodiester bond in the DNA backbone and under mildly acidic conditions, such as 1.5% trifluoroacetic acid (TFA) or 3-HPA/1% TFA MALDI-TOF matrix solution, results in the introduction of one or more intra-strand breaks in the DNA molecule. Methods for the preparation of diisopropylsilyl-linked oligonucleotide precursors and analogs are known to those of skill in the art (see e.g., Saha et al. (1993) J. Org. Chem. 58:7827-7831). These oligonucleotide analogs may be readily prepared using solid state oligonucleotide synthesis methods using diisopropylsilyl derivatized deoxyribonucleosides.

Modification of nucleic acids

A. Mass modification

- In certain embodiments, nucleic acids modified at positions other than the 3'- or 5'- terminus can be used. Modification of the sugar moiety of a nucleotide at positions other than the 3' and 5' position is possible through conventional methods. Also, nucleic acid bases can be modified, e.g., as described in F. Eckstein, ed., "Oligonucleotides and Analogues: A Practical Approach," IRL Press (1991). Such a linker arm can be modified to include a thiol moiety. Alternatively, backbone-modified nucleic acids can be used so that the thiol group can be attached to the nitrogen center provided by the modified phosphate backbone.
- In preferred embodiments, modification of a nucleic acid, such as described above, does not substantially impair the ability of the nucleic acid molecule to hybridize to its complement. Thus, any modification should preferably avoid substantially modifying the functionalities of the nucleic acid that are responsible for Watson-Crick base pairing. The

-35-

nucleic acid can be modified such that a non-terminal thiol group is present, and the nucleic acid, when immobilized to the support, is capable of self-complementary base pairing to form a "hairpin" structure having a duplex region.

5 **B. Other modifying RNA analogs**

In practicing the methods described herein, a set (or sets) of nested base-specific chain terminated RNA transcripts is (are) generated during transcription by the incorporation of a modified base-specific chain terminating ribonucleotide analog (or specific cleavage by RNases in a
10 modified Maxam-Gilbert strategy). Any ribonucleoside triphosphate analog that results in the sequence-specific arrest of transcription elongation upon incorporation into an RNA molecule by an RNA polymerase may be used in the methods herein. Presently preferred ribonucleotide analogs are 3'-deoxyribonucleotides, result in base-
15 specific termination of transcription (e.g., see Axelrod et al. (1985) Biochemistry 24:5716-5723; Tyagarajan et al. (1985) Biochemistry 30:10920-10924).

In certain embodiments, in addition to a base-specific chain terminating ribonucleoside triphosphate analog, additional ribonucleotide
20 analogs can be added to reduce the secondary structure of the resulting RNA transcript and/or to prevent premature transcription termination or pausing or RNA cleavage. For example, the incorporation of riboinosine using inosine 5'-triphosphate is known to reduce the secondary structure of RNA products. In the presence of a dinucleotide guanine initiator,
25 inosine 5'-triphosphate can effectively substitute for GTP in in vitro transcription reactions (e.g., Axelrod et al. (1985) Biochemistry 24:5716-5723).

In addition, modified ribonucleotide analogs may be added to the transcription mixture to increase the efficiency of transcriptional

-36-

termination and/or transcript release to promote and facilitate the rate enzyme turnover. For example, the addition of 4-thio UTP, 5-bromo UTP, 5-iodo CTP alter the hydrogen bonding of the nucleic acid facilitating, at least with some RNA polymerases, transcriptional termination and
5 transcript release.

Solid supports and substrates

Examples of insoluble supports and substrates for use herein include, but are not limited to: beads, such as silica gel, controlled pore glass, magnetic beads, dextran and agarose (Sephadex and Sepharose)
10 beads, cellulose beads and others, capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold, silver, aluminum, silicon and copper), plastic materials including multiwell plates or membranes (e.g., of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), wafers, combs, pins (e.g., arrays of pins
15 suitable for combinatorial synthesis or analysis) or beads in pits of flat surfaces such as wafers (e.g., silicon wafers), with or without filter plates.

Mass spectrometry

Once transcription is complete, the RNA transcripts can be
20 analyzed by any of a variety of means including, for example, spectrometric techniques such as UV/VIS, IR, fluorescence, chemiluminescence, or NMR spectroscopy, mass spectrometry, gel electrophoresis and other methods known in the art, or combinations thereof. Mass spectrometry is preferred herein. Mass spectrometer
25 formats include the ionization (I) techniques, such as matrix assisted laser desorption (MALDI), continuous or pulsed electrospray (ESI) and related methods (e.g. ionspray or Thermospray), and massive cluster impact (MCI); these ion sources can be matched with detection formats including linear or reflectron time-of-flight (TOF), single or multiple

-37-

quadruple, single or multiple magnetic sector, Fourier Transform ion cyclotron resonance (FTICR), ion trap, and combinations thereof to yield a hybrid detector (e.g., ion-trap/time-of-flight). For ionization, numerous matrix/wavelength combinations (MALDI) or solvent combinations (ESI) can be employed.

Preparation of DNA arrays

In preferred embodiments, nucleic acid promoter-containing probes, prior to or subsequent to hybridization to target nucleic acids, are immobilized at to the surface of a solid support in an array format.

10 Particularly suitable methods for forming these DNA arrays are those described herein and U.S. application Serial Nos. 08/748,053, 08/787,639 and 08/786,988. The subject matter of each pending U.S. applications is herein incorporated by reference in its entirety and is summarized as follows.

15 Figure 1 illustrates one system for preparing arrays of sample material for analysis by a diagnostic tool. Fig. 1 depicts a system 10 that includes a data processor 12, a motion controller 14, a robotic arm assembly 16, a monitor element 18A, a central processing unit 18B, a microliter plate of source material 20, a stage housing 22, a robotic arm 24, a stage 26, a pressure controller 28, a conduit 30, a mounting assembly 32, a pin assembly 38, and substrate elements 34. In the view shown by Fig. 1, it is also illustrated that the robotic assembly 16 can include a moveable mount element 40 and a horizontal slide groove 42. The robotic arm 24 can optionally pivot about a pin 36 to increase the travel range of the arm 24 so that arm 24 can disposes the pin assembly 25 38 above the source plate 20.

The data processor 12 depicted in Fig. 1 can be a conventional digital data processing system such as an IBM PC compatible computer system that is suitable for processing data and for executing program

-38-

instructions that will provide information for controlling the movement and operation of the robotic assembly 16. The data processor unit 12 can be any type of system suitable for processing a program of instructions signals that will operate the robotic assembly that is

5 integrated into the robotic housing 16. Optionally the data processor 12 can be a micro-controlled assembly that is integrated into robotic housing 16. In further alternative embodiments, the system 10 need not be programmable and can be a single board computer having a firmware memory for storing instructions for operating the robotic assembly 16.

10 In the embodiment depicted in Fig. 1, there is a controller 14 that electronically couples between the data processor 12 and the robotic assembly 16. The depicted controller 14 is a motion controller that drives the motor elements of the robotic assembly 16 for positioning the robotic arm 24 at a selected location. Additionally, the controller 14 can

15 provide instructions to the robotic assembly 16 to direct the pressure controller 28 to control the volume of fluid ejected from the individual pin elements of the depicted pin assembly 38. The design and construction of the depicted motion controller 14 follows from principles well known in the art of electrical engineering, and any controller element suitable for

20 driving the robotic assembly 16 can be practiced without departing from the scope thereof.

The robotic assembly 16 depicted in Fig. 1 electronically couples to the controller 14. The depicted robotic assembly 16 is a gantry system that includes an XY table for moving the robotic arm about a XY

25 plane, and further includes a Z axis actuator for moving the robotic arm orthogonally to that XY plane. The robotic assembly 16 depicted in Fig. 1 includes an arm 24 that mounts to the XY stage which moves the arm within a plane defined by the XY access. In the depicted embodiment, the XY table is mounted to the Z actuator to move the entire table along

-39-

the Z axis orthogonal to the XY plane. In this way, the robotic assembly provides three degrees of freedom that allows the pin assembly 38 to be disposed to any location above the substrates 34 and the source plate 20 which are shown in Fig. 1 as sitting on the stage 26 mounted to the robotic assembly 16.

The depicted robotic assembly 16 follows from principles well known in the art of electrical engineering and is just one example of a robotic assembly suitable for moving a pin assembly to locations adjacent a substrate and source plate such as the depicted substrate 34.

10 Accordingly, alternative robotic systems can be practiced following the descriptions herein without departing from the scope thereof.

Fig. 1 depicts an embodiment of a robotic assembly 16 that includes a pressure controller 28 that connects via a conduit 30 to the mount 32 that connects to the pin assembly 38. In this embodiment the mount 32 has an interior channel for fluidic coupling of the conduit 30 to the pin assembly 38. Accordingly, the pressure controller 28 is fluidically coupled by the conduit 30 and the mount 32 to the pin assembly 38. In this way the controller 14 can send signals to the pressure controller 28 to control selectively a fluid pressure delivered to the pin assembly 38.

20 Fig. 2 depicts one embodiment of a pin assembly 50 suitable for practice with the system depicted in Fig. 1 which includes the pressure controller 28. In the depicted embodiment, the pin assembly 50 includes a housing formed from an upper portion 52 and a lower portion 54 that are joined together by the crews 56A and 56B to define an interior chamber volume 58. Fig. 2 further depicts that to fluidically seal the interior chamber volume 58 the housing can include a seal element depicted in Fig. 2 as an O-ring gasket 60 that sites between the upper block and the lower block 54 and surrounds completely the perimeter of the interior chamber volume 58. Fig. 2 further depicts that the pin

-40-

assembly 50 includes a plurality of vesicles 62A-62D, each of which include an axial bore extending therethrough to form the depicted holding chambers 64A-64D. Each of the depicted vesicles extends through a respective aperture 68A-68D disposed within the lower block 54 of the
5 housing.

As further shown in the depicted embodiment, each of the vesicles 62A-62D has an upper flange portion that sits against a seal element 70A-70D to form a fluid-tight seal between the vesicle and the lower block 54 to prevent fluid from passing through the apertures 68A-68D.
10 To keep the seal tight, the depicted pin assembly 50 further includes a set of biasing elements 74A-74D depicted in Fig. 2 as springs which, in the depicted embodiments, are in a compressed state to force the flange element of the vesicles 62A-62D against their respective seal elements 70A-70D. As shown in Fig. 2, the biasing elements 74A-74D extend
15 between the vesicles and the upper block 52. Each of the springs 74A-74D can be fixedly mounted to a mounting pad 76A-76D where the spring elements can attach to the upper block 52. The upper block 52 further includes an aperture 78 depicted in Fig. 2 as a centrally disposed aperture that includes a threaded bore for receiving a swagelok 80 that
20 can be rotatably mounted within the aperture 78.

As further depicted in Fig. 2, the swagelok 80 attaches by a conduit to a valve 82 than can connect the swagelok 80 to a conduit 84 that can be coupled to a pressure source, or alternatively can couple the swagelok 80 to a conduit 86 that provides for venting of the interior
25 chamber 58. A central bore 88 extends through the swagelok 80 and couples to the tubing element which further connects to the valve 82 to thereby fluidically and selectively couple the interior chamber volume 58 to either a pressure source, or a venting outlet.

-41-

The pin assembly 50 described above and depicted in Fig. 2 disposed above a substrate element 90 that includes a plurality of wells 92 that are etched into the upper surface of the substrate 90. As illustrated by Fig. 2, the pitch of the vesicles 62A-62D is such that each vesicle is spaced from the adjacent vesicles by a distance that is an integral multiple of the pitch distance between wells 92 etched into the upper surface of the substrate 90. As will be seen from the following description, this spacing facilitates the parallel dispensing of fluid, such that fluid can be dispensed into a plurality of wells in a single operation.

Each of the vesicles can be made from stainless steel, silica, polymeric material or any other material suitable for holding fluid sample. In one example, 16 vesicles are employed in the assembly, which are made of hardened beryllium copper, gold plated over nickel plate. They are 43.2 mm long and the shaft of the vesicle is graduated to 0.46 mm outer diameter with a concave tip. Such a pin was chosen since the pointing accuracy can be approximately 501 micrometers. Any suitable pin style can be employed for the device, including but not limited to flat, star-shaped, concave, pointed solid, pointed semi-hollow, angled on one or both sides, or other such geometries.

Fig. 3 shows from a side perspective the lower block 54 of the pin assembly 50 depicted in Fig. 2. Fig. 3 shows approximate dimensions for one pin assembly. As shown, the lower block 54 has a bottom plate 98 and a surrounding shoulder 100. The bottom plate 98 is approximately 3 mm in thickness and the shoulder 100 is approximately 5 mm in thickness.

Fig. 4 shows from an overhead perspective the general structure and dimensions for one lower block 54 suitable for use with the pin assembly for use with the pin assembly 50 shown in Fig. 2. As shown in Fig. 4, the lower block 54 includes a four-by-four matrix of apertures

-42-

68 to provide 16 apertures each suitable for receiving a vesicle. As described above with reference to Fig. 2, the spacing between the aperture 68 is typically an integral multiple of the distance between wells on a substrate surface as well as the wells of a source plate.

- 5 Accordingly, a pin assembly having the lower block 54 as depicted in Fig. 4 can dispense fluid in up to 16 wells simultaneously. Fig. 4 also shows general dimensions of one lower block 54 such that each side of block 54 is generally 22 mm in length and the pitch between aperture 68 is approximately 4.5 mm. Such a pitch is suitable for use with a
- 10 substrate where fluid is to be dispensed at locations approximately 500 μm apart, as exemplified by the substrate 90 of Fig. 2. Fig. 4 also shows that the lower block 54 can include an optional O-ring groove 94 adapted for receiving an O-ring seal element, such as the seal element 60 depicted in Fig. 2. It is understood that such a groove element 94 can
- 15 enhance and improve the fluid seal formed by the seal element 60.

- The pinblock can be manufactured of stainless steel as this material can be drilled accurately to about $+25 \mu\text{m}$, but a variety of probe materials can also be used, such as G10 laminate, PMMA or other suitable material. The pin block can contain any number of apertures and
- 20 is shown with 16 receptacles which hold the 16 pins in place. To increase the pointing accuracy of each pin, an optional alignment plate can be placed below the block so that about 6 mm of the pin tip is left exposed to enable dipping into the wells of a microtiter plate. The layout of the probes in the depicted tool is designed to coordinate with a 384-
- 25 well microtiter plate, thus the center-to-center spacing of the probes in 4.5 mm. An array of 4 x 4 probes was chosen since it would produce an array that would fit in less than one square inch, which is the travel range of an xy stage of a MALDI TOF MS employed by the assignee.

-43-

The pintool assembly is completed with a stainless steel cover on the top side of the device which is then attached onto the Z-arm of the robot.

With reference to Figs 1 and 2, the robotic assembly 16 employs a pin tool assembly 38 that is configured similarly as the pin tool assembly 50 depicted in Figure 2. The pressure controller 28 selectively controls the pressure within chamber 58. With this embodiment, a control program operates on the data processor 12 to control the robotic assembly 16 in a way that the assembly 16 prints an array of elements on the substrates 34. In a first step, (see, also Figures 2 and 5), the program directs the robotic assembly 16 to move the pin assembly 38 to be disposed above the source plate 20. The robotic assembly 16 will then dip the pin assembly into the source plate 20 which can be a 384 well DNA source plate. As shown in Fig. 4 the pin assembly can include 16 different pins such that the pin assembly 50 will dip 16 pins into different 16 wells of the 384 well DNA source plate 20. Next the data processor 12 will direct the motion controller 14 to operate the robotic assembly 16 to move the pin assembly to a position above the surface of the substrate 34. The substrate 34 can be any substrate suitable for receiving a sample of material and can be formed of silicon, plastic, metal, or any other such suitable material. Optionally the substrate will have a flat surface, but can alternatively include a pitted surface, a surface etched with wells or any other suitable surface typography and further containing beads on the surface of each well. The program operating on data processor 12 can then direct the robotic assembly, through the motion controller 14, to direct the pressure controller 28 to generate a positive pressure within the interior chamber volume 58. In this practice, the positive interior pressure will force fluid from the holding chambers of vesicles 62 to eject fluid from the vesicles and into a respective well 92 of the substrate 90.

-44-

The program operating on data processor 12 can also direct the controller 14 to control the pressure controller 28 to control filling the holding chambers with source material from the source plate 20. The pressure controller 28 can generate a negative pressure within the interior chamber volume 58 of the pin assembly. This will cause fluid to be drawn up into the holding chambers of the vesicles 62A-62D. The pressure controller 28 can regulate the pressure either by open-loop or closed-loop control to avoid having fluid overdrawn through the holding chambers and spilled into the interior chamber volume 58. Loop control systems for controlling pressure are well known in the art and any suitable controller can be employed. Such spillage could cause cross-contamination, particularly if the source material drawn from the source plate 20 varies from well to well.

In an alternative embodiment, each of the holding chambers 64A-64D is sufficiently small to allow the chambers to be filled by capillary action. In such a practice, the pin assembly can include an array of narrow bore needles, such as stainless steel needles, that extend through the apertures of the lower block 54. The needles that are dipped into source solutions will be filled by capillary action. In one practice, the length of capillary which is to be filled at atmospheric pressure is determined approximately by:

$$H = \frac{2\gamma}{PGR}$$

where H equals Height, gamma equals surface tension, P equals solution density, G equals gravitational force and R equals needle radius. Thus the volume of fluid held by each vesicle can be controlled by selecting the dimensions of the interior bore. It is understood that at room temperature water will fill a 15 cm length of 100 μ m radius capillary. Thus, a short bore nanoliter volume needle will fill to full capacity, but should not overflow because the capillary force is understood to be too

-45-

small to form a meniscus at the top of the needle orifice. This prevents cross-contamination due to spillage. In one embodiment, the vesicles of the pin assembly can be provided with different sized interior chambers for holding and dispensing different volumes of fluid.

5 In an alternative practice, to decrease the volume of liquid that is drawn into the holding chambers of the vesicles, a small positive pressure can be provided within the interior chamber volume 58 by the pressure controller 28. The downward force created by the positive pressure can be used to counter the upward capillary force. In this way,
10 the volume of fluid that is drawn by capillary force into the holding chambers of the vesicles can be controlled.

Fig. 5B shows that fluid within the holding chambers of the needle can be dispensed by a small positive pressure introduced through the central bore 88 extending through a swagelok 80. By regulating the
15 pressure pulse that is introduced into the interior chamber volume 58, fluid can be ejected either as a spray or by droplet formation at the needle tip. It is understood that the rate of dispensing, droplet versus spray, depends in part upon the pressure applied by the pressure controller 28. In one practice, pressure is applied in the range of
20 between 10 and 1,000 Torr of atmospheric pressure.

To this end the data processor 12 can run a computer program that controls and regulates the volume of fluid dispensed. The program can direct the controller 28 to eject a defined volume of fluid, either by generating a spray or by forming a drop that sits at the end of the
25 vesicle, and can be contacted with the substrate surface for dispensing the fluid thereto.

Figures 5C and 5D show the earlier steps shown in Figs. 5A-5B can again be performed, this time at a position on the substrate surface that is offset from the earlier position. In the depicted process, the pin

-46-

tool is offset by a distance equal to the distance between two wells 92. Other offset printing techniques can be employed.

Several advantages of the pin assembly depicted in Fig. 2 are achieved. For example, rinsing between dispensing events is straight-
5 forward, requiring only single or multiple pin fillings and emptying events with a rinse solution. Moreover, since all holding chambers fill to full capacity, the accuracy of the volumes dispensed varies only according to needle inner dimensions which can be carefully controlled during pin
10 production. Further the device is cost effective, with the greatest expense attributed to the needles, however because no contact with a surface is required, the needles are exposed to little physical strain or stress, making replacement rare and providing long life.

Alternatively, deposition of sample material onto solid support surface can include techniques that employ pin tool assemblies that have
15 solid pin elements extending from a block wherein a robotic assembly dips the solid pin elements of the pin assembly into a source of sample material to wet the distal ends of the pins with the sample materials. Subsequently the robotic assembly can move the pin assembly to a location above the substrate and then lower the pin assembly against the
20 surface of the substrate to contact the individual wetted pins against the surface for spotting material of the substrate surface.

Pin tool assemblies that have solid pins with a promoter-containing nucleic acid immobilized thereon can be dipped into wells containing a reaction mixture for hybridization and/or generation of RNA transcript.
25 The liquid with the transcripts can then be deposited in an array as discussed above, for example by using solid pins of a pin tool assembly to spot sample on the substrate surface.

Figures 6A and 6B depict another alternative system for dispensing material on or to the surface of the substrate. In particular, Figure 6A

-47-

depicts a jet printing device 110 which includes a capillary element 112, a transducer element 114 and orifice (not shown) 118, a fluid conduit 122, and a mount 124 connecting to a robotic arm assembly, such as the robotic arm 24 depicted in Figure 1. As further shown in Figure 6A
5 the jet assembly 110 is suitable for ejecting from the orifice 118 a series of drops 120 of a sample material for dispensing sample material onto the surface 128.

The capillary 112 of the jet assembly 110 can be a glass capillary, a plastic capillary, or any other suitable housing that can carry a fluid
10 sample and that will allow the fluid sample to be ejected by the action of a transducer element, such as the transducer element 114. The transducer element 114 depicted in Figure 6A is a piezo electric transducer element which forms around the parameter of the capillary 112 and can transform an electrical pulse received from the pulse
15 generator within a robotic assembly 16 to cause fluid to eject from the orifice 118 of the capillary 112. One such jet assembly having a piezoelectric transducer element is manufactured by MicroFab Technology, Inc., of Germany. Any jet assembly, however, that is suitable for dispensing defined and controlled the volumes of fluid can be
20 used herein including those that use piezoelectric transducers, electric transducers, electrorestrictive transducers, magnetorestrictive transducers, electromechanical transducers, or any other suitable transducer element. In the depicted embodiment, the capillary 112 has a fluid conduit 122 for receiving fluid material. In an optional embodiment,
25 fluid can be drawn into the capillary by action of a vacuum pressure that will draw fluid through the orifice 118 when the orifice 118 is submerged in a source of fluid material. Other embodiments of the jet assembly 110 can be practiced with the invention without departing from the scope thereof.

-48-

Figure 6B illustrates a further alternative assembly suitable for p being carried on the robotic arm of a robotic assembly, such as the assembly 18 depicted in Figure 1. Figure 6B illustrates four jet assemblies connected together, 130A-130D. Similar to the pin assembly
5 In Figure 2, the jet assembly depicted in Figure 6B can be employed for the parallel dispensing of fluid material. Each of the jet assemblies 130A-130D can be operated independently of the others, allowing the selective dispensing of fluid from select ones of the jet assemblies. Moreover, each of the jet assemblies 130A-130D can be independently controlled to
10 select the volume of fluid that is dispensed from each respected one of the assembly 130A-130D. Other modifications and alterations can be made to the assembly depicted in Figure 6B without departing from the scope of the invention.

Arrays can be formed on a substrate surface according to any of
15 the techniques discussed above. The sample arrays are then analyzed by mass spectrometry to collect spectra data that is representative of the composition of the samples in the array. It is understood that the above methods provide processes that allow for rapidly dispensing definite and controlled volumes of analyte material. In particular these processes
20 allow for dispensing sub to low nanoliter volumes of fluid. These low volume deposition techniques generate sample arrays well suited for analysis by mass spectrometry. For example, the low volumes yield reproducibility of spot characteristics, such as evaporation rates and reduced dependence on atmospheric conditions such as ambient
25 temperature and light.

Continuing with the example shown in Fig. 5, the arrays can be prepared by loading oligonucleotides (0.1-50 ng/ μ l) of different sequences or concentrations into the wells of a 96 well microtiter source plate 20; the first well can be reserved for holding a matrix solution. A

-49-

substrate 34, such as a pitted silicon chip substrate, can be placed on the stage 26 of the robotics assembly 16 and can be aligned manually to orient the matrix of wells about a set of reference axes. The control program executing on the data processor 12 can receive the coordinates
5 of the first well of the source plate 20. The robotic arm 24 can dip the pin assembly 38 into source plate 20 such that each of the 16 pins is dipped into one of the wells. Each vesicle can fill by capillary action so that the full volume of the holding chamber contains fluid. Optionally, the program executing on the data processor 12 can direct the pressure
10 controller to fill the interior chamber 58 of the pin assembly 38 with a positive bias pressure that will counteract, in part, the force of the capillary action to limit or reduce the volume of fluid that is drawn into the holding chamber.

Optionally, the pin assembly 38 can be dipped into the same
15 wells of the source plate 20 and spotted on a second target substrate. This cycle can be repeated on as many target substrates as desired. Next the robotic arm 24 can dip the pin assembly 38 in a washing solution, and then dip the pin assembly into 16 different wells of the source plate 20, and spot onto the substrate target offset a distance
20 from the initial set of 16 spots. Again this can be repeated for as many target substrates as desired. The entire cycle can be repeated to make a 2x2 array from each vesicle to produce an 8x8 array of spots (2x2 elements/vesicle x 16 vesicles = 64 total elements spotted). Any process suitable for forming arrays can be employed.

25 Oligonucleotides of different sequences or concentrations can be loaded into the wells of up to three different 384-well microtiter source plates; one set of 16 wells can be reserved for matrix solution. The wells of two plates are filled with washing solution. Five microtiter plates can be loaded onto the stage of the robotic assembly 16. A

-50-

plurality of target substrates can be placed abutting an optional set of banking or registration pins disposed on the stage 26 and provided for aligning the target substrates along a set of reference axes. If the matrix and oligonucleotide are not pre-mixed, the pin assembly can be employed
5 to first spot matrix solution on all desired target substrates. In a subsequent step the oligonucleotide solution can be spotted in the same pattern as the matrix material to re-dissolve the matrix. Alternatively, a sample array can be made by placing the oligonucleotide solution on the wafer first, followed by the matrix solution, or by pre-mixing the matrix
10 and oligonucleotide solutions.

After depositing the sample arrays onto the surface of the substrate, the arrays can be analyzed using any of a variety of means (e.g., spectrometric techniques, such as UV/VIS, IR, fluorescence, chemiluminescence, NMR spectrometry or mass spectrometry. For
15 example, subsequent to either dispensing process, sample loaded substrates can be placed onto a MALDI-TOF source plate and held there with a set of beveled screw mounted polycarbonate supports. In one practice, the plate can be transferred on the end of a probe to be held onto a 1 μ m resolution, 1" travel xy stage (Newport) in the source region
20 of a time-of-flight mass spectrometer. Any suitable mass spectrometry tool can be employed.

Preferred mass spectrometer formats for use with the arrays described herein include ionization (I) techniques including but not limited to matrix assisted laser desorption (MALDI), continuous or pulsed
25 electrospray (ESI) and related methods (e.g., ionspray or Thermospray), or massive cluster impact (MCI); those ion sources can be matched with detection formats including linear or non-linear reflectron time-of-flight (TOF), single or multiple quadrupole, single or multiple magnetic sector, Fourier Transform ion cyclotron resonance (FTICR), ion trap, and

-51-

combinations thereof (e.g., ion-trap/time-of-flight). For ionization, numerous matrix/wavelength combinations (MALDI) or solvent combinations (ESI) can be employed. Subattomole levels of protein have been detected for example, using ESI (Valaskovic, G. A. et al., {1996} Science 273: 1199-1202) or MALDI (Li, L. et al., {1996} J. Am. Chem. Soc 118: 1662-1663) mass spectrometry.

Thus, it will be understood that in processes described herein a completely non-contact, high-pressure spray or partial-contact, low pressure droplet formation mode can be employed. In the latter, the only
10 contact that will occur is between the droplet and the walls of the well or a hydrophilic flat surface of the substrate 34. In neither practice need there be any contact between the needle tip and the surface.

Exemplary embodiment

In one preferred embodiment, a double-stranded nucleic acid
15 sequence encoding a promoter sequence is isolated from a natural source (e.g., bacteria, viruses, bacteriophages, plants or eukaryotic organisms) or assembled from synthetic sequences. A single-stranded region of at least a 5 nucleotides at the 3'-end of the coding strand using standard methods known to those of skill in the art (e.g., see Sambrook et al.,
20 {1989} Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory Press, New York). This single-stranded region is designed such that it is complementary to a region of the nucleic acid to be sequenced or to an sequence shared between the two nucleic acid molecules (e.g., a restriction endonuclease site).

25 The nucleic acid to be sequenced containing at least a partially single-stranded 3'-end is hybridized according to the conditions described herein, preferably high stringency such that 5 contiguous matching base pairs form, and known to those of skill in the art to the complementary sequences of the promoter-containing DNA. The nucleic acid to be

-52-

sequenced may be double-stranded or preferably single-stranded. The hybridization of the two nucleic acid molecules introduces one or more "nicks" in the hybrid at the junction(s) of the adjacent nucleic acid molecules. Nicks in the coding or non-coding strand, preferably the coding strand, can be ligated by the addition of an appropriate nucleic acid ligase prior to initiating transcription. Methods for ligating nucleic acids are well known to those of skill in the art (e.g., see Sambrook et al., (1989) Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory Press, New York) and DNA and RNA ligases are commercially available (e.g., Boehringer Mannheim, Indianapolis, IN).

Transcription is initiated from the promoter by the addition of the appropriate RNA polymerase in the presence of ribonucleoside triphosphates under conditions described herein and known elsewhere (e.g., In RNA Polymerase and the Regulation of Transcription, Reznikoff et al., eds. Elsevier, NY). In preferred embodiments, a selected base-specific chain terminating 3'-deoxyribonucleoside triphosphate and the transcription mixture also contains inosine 5'-triphosphate to reduce the secondary structure of the RNA product or modified ribonucleoside triphosphates, such as 4-thio UTP, 5-bromo UTP or 5'-iodo CTP to facilitate turnover of the RNA polymerase enzyme and thereby increase the amount of RNA transcript available for analysis.

In preferred embodiments, the nucleic acid promoter-containing probe is covalently immobilized on a silica support by functionalization of the support with an amino functionality (e.g., by derivatization of the support with a reagent such as 3-aminopropyl-triethoxysilane (Aldrich Chemical Co., Milwaukee, WI); see Figure 7). Other functionalized oxysilanes or orthosilicates can be used, and are commercially available (e.g., from Gelest, Inc., Tullytown, PA). For example, 3-mercaptopropyl-triethoxysilane can be used to functionalize a silicon surface with thiol

-53-

groups. The amino-functionalized silica can then be reacted with a heterobifunctional reagent such as N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB) (Pierce, Rockford, IL). Other homo- and hetero-bifunctional reagents which can be employed are available commercially,
5 e.g., from Pierce. Finally, a nucleic acid functionalized with a thiol group (e.g., at the 5'-terminus) is covalently bound to the derivatized silica support by reaction of the thiol functionality of the nucleic acid molecule with the iodoacetyl functionality of the support.

In certain embodiments, the nucleic acid can be reacted with the
10 cross-linking reagent to form a cross-linker/nucleic acid conjugate, which is then reacted with a functionalized support to provide an immobilized nucleic acid. Alternatively, the cross-linker can be combined with the nucleic acid and a functionalized solid support in one pot to provide substantially simultaneous reaction of the cross-linking reagent with the
15 nucleic acid and the solid support. In this embodiment, it will generally be necessary to use a heterobifunctional cross-linker, i.e., a cross-linker with two different reactive functionalities capable of selective reaction with each of the nucleic acid and the functionalized solid support.

Following the methods described herein, spatially-addressable
20 arrays of nucleic acids immobilized on insoluble supports that are suitable for sequencing of nucleic acids using RNA polymerase can be prepared. For example, the methods can be used to provide arrays of different nucleic acids immobilized on pins arranged in an array. In another embodiment, a photo-cleavable protecting group on the insoluble support
25 can be selectively cleaved (e.g., by photolithography) to provide portions of a surface activated for immobilization of a nucleic acid. For example, a silicon surface, modified by treatment with 3-mercaptopropyl-triethoxysilane to provide thiol groups, can be blocked with a photocleavable protecting group (for examples of photocleavable

-54-

protecting groups, see, e.g., PCT Publication WO 92/10092, or McCray et al., (1989) Ann. Rev. Biophys. Biophys. Chem. 18:239-270), and be selectively deblocked by irradiation of selected areas of the surface, e.g., by use of a photolithography mask. A nucleic acid promoter-containing probe modified to contain a thiol-reactive group can then be attached directly to the support, or, alternatively, a thiol-reactive cross-linking reagent can be reacted with the thiol-modified support, followed by (or substantially simultaneously with) reaction with a nucleic acid to provide immobilized nucleic acids. A nucleic acid base or sequence, once immobilized on a support according to the methods described herein, can be further modified according to known methods. For example, the nucleic acid molecule can be lengthened by performing solid-phase nucleic acid synthesis according to conventional techniques, including combinatorial technique

Preferably the nucleic acids are covalently bound to a surface of the insoluble support through at least one sulfur atom, i.e., the nucleic acids are covalently bound to the surface through a linker moiety which includes at least one sulfur atom. Such covalently bound nucleic acids are readily produced by the methods described herein. In preferred embodiments, the covalently bound nucleic acids are present on the surface of the insoluble support at a density of at least about 20 fmol/mm², more preferably at least about 75 fmol/mm², still more preferably at least about 100 fmol/mm², yet more preferably at least about 150 fmol/mm².

Applications

The methods of sequencing nucleic acids described herein may be used for a variety of end use applications. For example, the methods may be used for diagnostic applications for the identification of mutations, such as transitions, transversions, deletions, insertions and

-55-

the like. The methods may also be used to assist in the diagnosis of a number of genetic disorders, used for genetic screening or the determination of heredity or tissue or organ compatibility.

The method of sequencing may be used for diagnostic applications

- 5 to determined the presence of genetic alterations in a known target nucleic acid. For example, a region of the target nucleic acid can be amplified using standard methods, such as PCR or other amplification methods known to those of skill in the art (e.g., see Sambrook et al., (1989) Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory Press,
10 New York). The amplified nucleic acid can be denaturated and the strand to be sequenced (the noncoding strand, is isolated or may be used as a double-stranded molecule.

- The presence of genetic alterations in a known target nucleic is determined in a diagnostic application by isolating (or amplifying) a target
15 nucleic acid from a biological sample and hybridizing the target to a promoter containing capture probe to generate a promoter containing template. A nested set of RNA transcripts is generated and the sequence of the target nucleic acid determined and compared with a wildtype sequence to determine if a genetic alteration has occurred in the
20 target nucleic acid. The presence of a genetic alteration can be determined from the RNA transcripts by other means known in the art (see, e.g., U.S. Patent No. 5,605,798 and International PCT application No. WO 98/20019).

Methods of Identifying Transcriptional Terminator Sequences

- 25 Methods of identifying heretofore unknown transcriptional terminator and attenuator sequences are also provided. Transcriptional terminator sequences are responsible for the sequence specific cessation of RNA elongation and such terminator sequences from a variety of organisms have been reported. For example, bacterial rho-independent

-56-

terminator sequences have two inverted repeats separate by several base pairs followed by a 5-10 nt polyT stretch. Upon transcription through these sequences, the resulting mRNA transcript forms a RNA hairpin stem-loop secondary structure behind the RNA polymerase molecule that increases pausing of RNA polymerase and/or destabilizes the RNA polymerase-DNA template interaction resulting in termination of the transcript within the DNA polyT stretch. (e.g., see Wilson and von Hippel (19; Reynolds et al. (1992a) J. Mol. Biol. 224:53-63; Reynolds et al. (1992b) J. Mol. Biol. 224:31-51; Telesnitsky et al. (1989) Biochemistry 28:5210-5218; d'Aubenton Carafa et al. (1990) J. Mol. Biol. 216:835-858).

Bacterial rho-dependent terminators lack the traditional inverted repeat stem-loop structure and further require additional factors, such as rho protein, to halt transcription (e.g., see Schmidt et al. (1984) J. Mol. Biol. 259:15000-15002). Rho-dependent termination is believed to result in the premature termination in bacterial species upon uncoupling of transcription and translation.

By modifying the standard transcription conditions described herein, transcriptional terminator sequences, e.g., rho-dependent and rho-independent terminators, may be identified using mass spectrometric methods. In practicing the methods, a single-stranded region of the 3'-end of the nucleic acid to be sequenced is hybridized to a complementary sequence at the 3'-end of the coding strand a promoter-containing nucleic acid probe. In preferred embodiments, the promoter-containing nucleic acid is covalently coupled via the 5'-end of the noncoding strand or 3'-end of the coding strand to a solid support and, more preferably, is a 5'- or 3'-thiolated DNA linked at high densities to a aminosilane-treated solid support. The linkage may be in the absence or presence of a linker group and is preferably arranged in an array format.

-57-

Transcription is initiated in the absence or presence of modified RNA triphosphate analogs that increase the efficiency of RNA polymerase termination at such terminator sequences, such as 4-thio UTP, 5-bromo UTP or 5'-iodo CTP. The mass of the specifically terminated RNA transcripts can be detected by mass spectrometry where the observed mass of the RNA is indicative of the location of the terminator-dependent arrest of transcription. By comparing of the alignment of the sequence immediately preceding the site of transcriptional termination from several distinct genomic locations, heretofore unknown terminator sequences may be identified for different RNA polymerases.

In certain embodiments, nicks in one or more strand resulting from the hybridization of the nucleic acid to be sequenced may be ligated by the addition of an appropriate nucleic acid ligase prior to initiating transcription (i.e., adding a DNA or RNA ligase).

The present invention is further illustrated by the following Examples, which are intended merely to further illustrate and should not be construed as limiting. The entire contents of all the of the references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

EXAMPLE 1

High Density Attachment of nucleic Acids to Silicon Wafers

Materials and Methods

All reagents, unless otherwise noted, were obtained from Aldrich Chemical, Milwaukee, WI.

-59-

oligodeoxynucleotide upon cleavage of the disulfide. Various reduction methods were investigated to determine the optimal conditions. In one case, the disulfide-containing oligodeoxynucleotide (31.5 nmol, 0.5 mM) was incubated with dithiothreitol (DTT) (Pierce Chemical, Rockford, IL) (6.2 mmol, 100 mM) at pH 8.0 and 37°C. With the cleavage reaction essentially complete, the free thiol-containing oligodeoxynucleotide was isolated using a Chromaspin-10 column (Clontech, Palo Alto, CA) since DTT may compete in the subsequent reaction. Alternatively, tris-(2-carboxyethyl) phosphine (TCEP) (Pierce Chemical, Rockford, IL) has been used to cleave the disulfide. The disulfide-containing oligodeoxynucleotide (7.2 nmol, 0.36 mM) was incubated with TCEP in pH 4.5 buffer at 37°C. It is not necessary to isolate the product following the reaction since TCEP does not competitively react with the iodoacetamido functionality. Varying concentrations of TCEP were used for the cleavage reaction to determine the optimal conditions for the conjugation reaction.

Probe Coupling

To each wafer which had been derivatized to contain the iodoacetamido functionality as described above was added a 10 mM aqueous solution of the free-thiol containing oligodeoxynucleotide in 100 mM phosphate buffer, pH 8; the reaction was allowed to proceed for a minimum of five hours at room temperature in 100% relative humidity. Following the reaction, the oligodeoxynucleotide solution was removed, and the wafers were washed two times in 5 X SSC buffer (75 mM sodium citrate, 750 mM sodium chloride, pH 7) with 50% formamide (USB, Cleveland, OH) at 65°C for 1 hour each.

-60-

Radiochemical Determination of Probe Density

In order to determine the amount of DNA covalently attached to a surface or the amount of a complementary sequence hybridized, radiolabeled probes were employed. In cases where a 5'-disulfide-
5 containing oligodeoxynucleotide was to be immobilized, the 3'-terminus was radiolabeled using terminal transferase enzyme and a radiolabeled dideoxynucleoside triphosphate; in a standard reaction, 15 pmol (0.6 μ M) of the 5'-disulfide-containing oligodeoxynucleotide was incubated with 50 μ Ci (16.5 pmol, 0.66 μ M) of (α - 32 P) dideoxyadenosine-5'triphosphate
10 (ddATP) (Amersham, Arlington Height, IL) in the presence of 0.2 mM 2-mercaptoethanol. Upon the addition of 40 units of the terminal deoxynucleotidyl transferase enzyme (USB, Cleveland, OH), the reaction was allowed to proceed for one hour at 37°C. After this time, the reaction was stopped by immersion of the vial in 75°C water bath for ten
15 minutes, and the product was isolated using a Chromaspin-10 column (Clontech, Palo Alto, CA). Similarly, a 5'-disulfide-containing oligodeoxynucleotide was radiolabeled with 35 S.

In cases where a 3'-disulfide-containing oligodeoxynucleotide was to be immobilized, the 5'-terminus was radiolabeled using T4
20 polynucleotide kinase and a radiolabeled nucleoside triphosphate. For example, 15 pmol (0.6 μ M) of the 3'-disulfide-containing oligodeoxynucleotide was incubated with 50 μ Ci (16.5 pmol, 0.66 μ M) of (γ - 32 P) adenosine-5'triphosphate (ATP) (Amersham, Arlington Height, IL) in the presence of 50 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 10 mM 2-
25 mercaptoethanol. Following the addition of 40 units of T4 polynucleotide kinase, the reaction was allowed to proceed for 1 hour at 37°C. The reaction was stopped by immersion of the vial in a 75°C water bath for ten minutes; the product was then isolated using a Chromaspin-10 column (Clontech, Palo Alto, CA).

-61-

To determine the density of covalently immobilized probe, the disulfide-containing oligodeoxynucleotide of choice was added to a trace amount of the same species than had been radiolabeled as described above. The disulfide was cleaved, the probe was immobilized on

5 iodoacetamido-functionalized wafers, the wafers were washed, and then exposed to a phosphorimager screen (Molecular Dynamics, Sunnyvale, CA). For each different oligodeoxynucleotide used, reference spots were made on polystyrene in which the molar amount of oligodeoxynucleotide was known; these reference spots were exposed to the phosphorimager

10 screen as well. Upon scanning the screen, the quantity (in moles) of oligodeoxynucleotide bound to each chip was determined by comparing the counts to the specific activities of the references.

Hybridization and Efficiency

To a wafer that had been functionalized with an immobilized probe

15 was added a solution of a complementary sequence (10 μ M) in 1 M NaCl and TE buffer. The wafer and solution were heated to 75°C and allowed to cool to room temperature over 3 hours. After this time, the solution was removed, and the wafer was washed two times with TE buffer.

To determine the amount of oligonucleotide hybridized,

20 immobilization of the probe was first carried out as described above except that the probe was labeled with ^{35}S rather than ^{32}P . The density of immobilized probe was determined with the phosphorimager. Next, the same wafer was incubated in TE buffer, 1 M NaCl, and its complementary strand (10 μ M) which had been radiolabeled with ^{32}P .

25 Hybridization was carried out as previously described. Following a wash to remove non-specific binding, the wafer and reference were exposed to a phosphorimager screen with a piece of copper foil between the screen and the wafer. The copper foil serves to block the signal from ^{35}S , while allowing the ^{32}P signal to pass freely. The molar amount of hybridized

-62-

oligonucleotide is then determined, thus revealing the percent of covalently immobilized probe that is available for hybridization.

MALDI-TOF Mass Spectrometric Analysis

- As described above, wafers containing non-radiolabeled
- 5 immobilized oligodeoxynucleotide (name: TCUC; sequence: GAATTCGAGCTCGGTACCCGG; molecular weight; 6455 Da; SEQ ID NO. 1) were synthesized, and a complementary sequence (name: MJM6; sequence: CCGGGTACCGAGCTCGAATTC; molecular weight: 6415 Da; SEQ ID NO. 2) was hybridized. The wafers were washed in
- 10 50mM ammonium citrate buffer for cation exchange to remove sodium and potassium ions on the DNA backbone (Pieles, U. et al., (1993) Nucl. Acids Res. 21:3191-3196). A matrix solution of 3-hydroxypicolinic acid (3-HPA, 0.7 M in 50% acetonitrile, 10% ammonium citrate; Wu, K.J., et al. (1993) Rapid Commun. Mass Spectrom. 7:142-146) was spotted
- 15 onto the wafer and allowed to dry at ambient temperature. The wafers were attached directly to the sample probe of a Finnigan MAT (Bremen, Germany) Vision 2000 reflectron TOF mass spectrometer using a conducting tape. The reflectron possesses a 5 keV ion source and 20 keV post-acceleration; a nitrogen laser was employed; and all spectra
- 20 were taken in the positive ion mode.

Surface Chemistry

- Employing standard silicon dioxide modification chemistry, a silicon wafer was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amino groups on the surface. As shown in
- 25 Figure 7, the surface was then exposed to a heterobifunctional crosslinker resulting in iodoacetamido groups on the surface. It was possible to determine the optimal reaction time of this reaction in solution using TLC. The SIAB crosslinker was visualized under long wave ultraviolet light (302 nm) to reveal a spot with an R_f value of 0.58. 3-

-63-

aminopropyltriethoxysilane was not active under ultraviolet light, therefore, ninhydrin was used to reveal a purple spot indicating the presence of a primary amine at the baseline. A microscale reaction was run using a slight molar excess of SIAB in comparison to 3-aminopropyltriethoxysilane; TLC analysis after approximately one minute revealed a new spot visible under long wave ultraviolet light with an R_f value of 0.28. There was no evidence of a purple spot upon spraying with ninhydrin, thus all the 3-aminopropyltriethoxysilane starting material had been consumed in the reaction. UV light also revealed the excess SIAB which remained following the reaction. From these results, it was determined the reaction is complete after approximately one minute. In all cases, the iodoacetamido-functionalized wafers were used immediately to minimize hydrolysis of the labile iodoacetamido-functionality. Additionally, all further wafer manipulations were performed in the dark since the iodoacetamido-functionality is light sensitive.

Disulfide reduction of the modified oligonucleotide was monitored by observing a shift in retention time on reverse-phase FPLC. It was determined that after five hours in the presence of DTT (100 mM) or TCEP (10 mM), the disulfide was fully reduced to a free thiol. If the DTT reaction was allowed to proceed for a longer time, an oligonucleotide dimer formed in which pairs of free thiols had reacted. Such dimerization was also observed when the DTT was removed following the completion of the cleavage reaction. This dimerization was not observed when TCEP was employed as the cleavage reagent since this reaction is performed at pH 4.5, thus the free thiols were fully protonated inhibiting dimerization.

Immediately following disulfide cleavage, the modified oligonucleotide was incubated with the iodoacetamido-functionalized

-64-

wafers. To ensure complete thiol deprotonation, the coupling reaction was performed at pH 8.0. The probe surface density achieved by this chemistry of silicon wafers was analyzed using radiolabeled probes and a phosphorimager. The probe surface density was also monitored as a function of the TCEP concentration used in the disulfide cleavage reaction (Figure 8). Using 10 mM TCEP to cleave the disulfide and the other reaction conditions described above, it was possible to reproducibly yield a surface density of 250 fmol per square mm of surface. Identical experiments as described above were performed except that the oligonucleotide probe lacked a thiol modification; surface densities of less than 5 fmol per square mm of surface proved that non-specific binding is minimal and that probe coupling most likely occurred as proposed in Figure 7.

Hybridization

After attaching ^{35}S -labeled probes to the surface of wafers and determining conjugation density as described above, hybridization of ^{32}P -labeled oligonucleotides was carried out; hybridization efficiency and density were determined using the phosphorimager and copper foil. It was determined experimentally that copper foil blocks 98.4% of an ^{35}S signal, while fully allowing a ^{32}P signal to be detected. The complementary sequence reproducibly hybridized to yield 105 fmol per square mm of surface; this corresponds to approximately 40% of the conjugated probes available for hybridization. Similarly, a non-complementary sequence was employed in this scheme yielding less than 5 fmol per square mm of surface in non-specific binding.

It is hypothesized that steric interference between the tightly packed oligonucleotide on the flat surface inhibits hybridization efficiencies higher than 40%. With this in mind, a spacer molecule was incorporated between the terminus of the hybridizing region of the

-65-

oligonucleotide and the support. The chosen spacers were a series of poly dT sequences ranging in length from 3 to 25. Upon examination of these samples with radiolabels and the phosphorimager, it was determined that 40% was still the maximum hybridization that could be
5 achieved.

MALDI-TOF MS Analysis

Wafers were functionalized with probes, complementary sequences were hybridized, and the samples were analyzed under standard MALDI conditions as described above. Analysis revealed that
10 only the annealed strand (MJM6) was observed in the mass spectrum with an experimental mass-to-charge ratio of 6415.4; the theoretical mass-to-charge ratio is 6415 (Figure 9). Since there was no signal at a mass-to-charge ratio of 6455, it was determined that the wafer-conjugated strand (TCUC) was not desorbed thus the iodoacetamido
15 linkage was stable enough to withstand the laser and remain intact. There was an additional signal observed at a mass-to-charge ratio of 6262.0. This signal results from a depurination of guanosines since it is known that DNA is susceptible to the loss of purine bases during the MALDI process, (Nordoff, E., et al., (1992) Rapid Commun. Mass
20 Spectrom. 6:771-776). The sample crystals on the wafer were not homogeneously distributed, thus it was necessary to hunt for a good spot. Because of this non-homogeneity, the mass resolution varied, but it generally ranged from 200-300 for the desorbed oligonucleotide in the mass spectra. In one set of experiments, non-complementary sequences
25 were hybridized to the wafer; following a wash as previously described, analysis by MALDI-TOF MS revealed that minimal non-specific annealing had taken place since no signal was detected.

-66-

EXAMPLE 2**Preparation of DNA arrays using serial and parallel dispensing tools**

Robot-driven serial and parallel pL-nL dispensing tools were used to generate $10\text{-}10^3$ element DNA arrays on <1 " square chips with flat or
5 geometrically altered (e.g. with wells) surfaces for matrix assisted laser desorption ionization mass spectrometry analysis. In the former, a 'piezoelectric pipette' ($70\text{ }\mu\text{m}$ id capillary) dispenses single or multiple ~ 0.2 nL droplets of matrix, and then analyte, onto the chip; spectra from as low as 0.2 fmol of a 36-mer DNA have been acquired using this
10 procedure. Despite the fast (<5 sec) evaporation, micro-crystals of 3-hydroxypicolinic acid matrix containing the analyte are routinely produced resulting in higher reproducibility than routinely obtained with larger volume preparations; all of 100 five fmol spots of a 23-mer in $800\text{ }\mu\text{m}$ wells yielded easily interpreted mass spectra, with 99/100 parent ion
15 signals having signal to noise ratio of >5 . In a second approach, probes from 384 well microtiter plate are dispensed 16 at a time into chip wells or onto flat surfaces using an array of spring loaded pins which transfer ~ 20 nL to the chip by surface contact; MS analysis of array elements deposited with the parallel method are comparable in terms of sensitivity
20 and resolution to those made with the serial method.

Description of the piezoelectric serial dispenser

The experimental system developed from a system purchased from Microdrop GmbH, Norderstedt Germany and can include a piezoelectric element driver which sends a pulsed signal to a piezoelectric element
25 bonded to and surrounding a glass capillary which holds the solution to be dispensed; a pressure transducer to load (by negative pressure) or empty (by positive pressure) the capillary; a robotic xyz stage and robot driver to maneuver the capillary for loading, unloading, dispensing, and cleaning, a stroboscope and driver pulsed at the frequency of the piezo

-67-

element to enable viewing of 'suspended' droplet characteristics;
separate stages for source and designation plates or sample targets (i.e.
Si chip); a camera mounted to the robotic arm to view loading to
designation plate; and a data station which controls the pressure unit,
5 xyz robot, and piezoelectric driver.

Description of the parallel dispenser

The robotic pintool includes 16 probes housed in a probe block
and mounted on an X Y, Z robotic stage. The robotic stage was a gantry
system which enables the placement of sample trays below the arms of
10 the robot. The gantry unit itself is composed of X and Y arms which
move 250 and 400 mm, respectively, guided by brushless linear servo
motors with positional feedback provided by linear optical encoders. A
lead screw driven Z axis (50 mm vertical travel) is mounted to the xy axis
slide of the gantry unit and is controlled by an in-line rotary servo motor
15 with positional feedback by a motor-mounted rotary optical encoder. The
work area of the system is equipped with a slide-out tooling plate that
holds five microtiter plates (most often, 2 plates of wash solution and 3
plates of sample for a maximum of 1152 different oligonucleotide
solutions) and up to ten 20x20 mm wafers. The wafers are placed
20 precisely in the plate against two banking pins and held secure by
vacuum. The entire system is enclosed in plexi-glass housing for safety
and mounted onto a steel support frame for thermal and vibrational
damping. Motion control is accomplished by employing a commercial
motion controller which was a 3-axis servo controller and is integrated to
25 a computer; programming code for specific applications is written as
needed.

Samples were dispensed with the serial system onto several
surfaces which served as targets in the MALDI TOF analysis including (1)
A flat stainless steel sample target as supplied for routine use in a

-68-

Thermo Bioanalysis Vision 2000; (2) the same design stainless steel target with micromachined nanopits; (3) flat silicon (Si) wafers; (4) polished flat Si wafers; (5) Si wafers with rough (3-6 μm features) pits; (6)(a) 12x12 or (b) 18x18 mm Si chips with (a) 10x10 (or (b) 18x18) arrays of chemically etched wells, each 800x800 μm on a side with depths ranging from 99-400 (or(b) 120) micrometer, pitch (a) 1.0 (or(b) 1.125) mm; (7) 15x15 mm Si chips with 28x28 arrays of chemically etched wells, each 450x450 micrometer on a side with depths ranging from 48-300 micrometer, pitch 0.5 mm; (8) flat polycarbonate or other plastics; (9) gold and other metals; (10) membranes; (11) plastic surfaces sputtered with gold or other conducting materials. The dispensed volume is controlled from 10^{-10} to 10^{-6} L by adjusting the number of droplets dispensed.

Sample Preparation and Dispensing

1. Serial

Oligonucleotides (0.1-50 ng/microliter of different sequence or concentrations were loaded into wells of a 96 well microtiter plate; the first well was reserved for matrix solution. A pitted chip (target 6a in MALDI targets' section) was placed on the stage and aligned manually. Into the (Windows-based) robot control software were entered the coordinates of the first well, the array size (ie number of spots in x and y) and spacing between elements, and the number of 0.2 nL drops per array element. The capillary was filled with $\sim 10 \mu\text{L}$ rinse H_2O , automatically moved in view of a strobe light-illuminated camera for checking tip integrity and cleanliness while in continuous pulse mode, and emptied. The capillary was then filled with matrix solution, again checked at the stroboscope, and then used to spot an array onto flat or pitted surfaces. For reproducibility studies in different MS modes, typically a 10x10 array of 0.2-20 nL droplets were dispensed. The

-69-

capillary was emptied by application of positive pressure, optionally rinsed with H₂O, and led to the source oligo plate where ~5 μ L of 0.05-2.0 μ M synthetic oligo were drawn. The capillary was then rastered in series over each of the matrix spots with 0.2-20 nL aqueous solution
5 added to each.

2. Parallel

Parallel Programs were written to control array making by offset printing; to make an array of 64 elements on 10 wafers, for example, the tool was dipped into 16 wells of a 384 well DNA source plate, moved to
10 the target (e.g. Si, plastic, metal), and the sample spotted by surface contact. The tool was then dipped into the same 16 wells and spotted on the second target; this cycle was repeated on all ten wafers. Next the tool was dipped in washing solution, then dipped into 16 different wells of the source plate, and spotted onto the target 2.25 mm offset
15 from the initial set of 16 spots; again this was repeated on all 10 wafers; the entire cycle was repeated to make a 2x2 array from each pin to produce an 8x8 array of spots (2x2 elements/pin X 16 pins = 64 total elements spotted).

To make arrays for MS analysis, oligonucleotides of different
20 sequences or concentrations were loaded into the wells of up to three different 384-well microtiter plates, one set of 16 wells was reserved for matrix solution. The wells of two plates were filled with washing solution. The five microtiter plates were loaded onto the slide-out tooling plate. Ten wafers were placed abutting the banking pins on the tooling
25 plate, and the vacuum turned on. In cases where matrix and oligonucleotide were not pre-mixed, the pintool was used to spot matrix solution first on all desired array elements of the ten wafers. For this example, a 16 x 16 array was created, thus the tool must spot each of the ten wafers 16 times, with an offset of 1.125 mm. Next, the

-70-

oligonucleotide solution was spotted in the same pattern to re-dissolve the matrix. Similarly, an array could be made by placing the oligonucleotide solution on the wafer first, followed by the matrix solution, or by pre-mixing the matrix and oligonucleotide solutions.

5 Mass spectrometry

Subsequent to either dispensing scheme, loaded chips were held onto a MALDI-TOF source plate with a set of beveled screw mounted polycarbonated supports. The plate was transferred on the end of a probe to be held onto a 1 μ m resolution, 1" travel xy stage (Newport) in
10 the source region of a time-of-flight mass spectrometer. The instrument, normally operated with 18-26 kV extraction, could be operated in linear or curved field reflectron mode, and in continuous or delayed extraction mode.

Serial dispensing with the piezoelectric pipette

15 While delivery of a saturated 3HPA solution can result in tip clogging as the solvent at the capillary-air interface evaporates, pre-mixing DNA and matrix sufficiently dilutes the matrix such that it remains in solution while stable sprays which could be maintained until the capillary was emptied were obtained; with 1:1 diluted (in H₂O) matrix
20 solution, continuous spraying for >> 10 minutes was possible. Turning off the piezo element so that the capillary sat inactive for >5 minutes, and reactivating the piezo element also did not result in a clogged capillary.

Initial experiments using stainless steel sample targets as provided
25 by Finnigan Vision 2000 MALDI-TOF system run in reflectron mode used a pre-mixed solution of the matrix and DNA prior to dispensing onto the sample target. In a single microtiter well, 50 μ L saturated matrix solution, 25 μ L of a 51 μ L solution of the 12-mer (ATCG)₃ (SEQ ID No. 3), and 25 μ L of a 51 μ L solution of the 28-mer (ATCG)₇ (SEQ ID No. 4) were

-71-

mixed. A set of 10x10 arrays of 0.6 μ L drops was dispensed directly onto a Finnigan Vision 2000 sample target disk; MALDI-TOF mass spectrum was obtained from a single array element which contained 750 attomoles of each of the two oligonucleotides. Interpretable mass spectra
5 has been obtained for DNA molecules as large as a 53-mer (350 amol loaded, not shown) using this method.

Mass spectra were also obtained from DNA molecules microdispensed into the wells of a silicon chip. Figure 10 shows a 12x12 mm silicon chip with 100 chemically etched wells; mask
10 dimensions and etch time were set such that fustum (i.e., inverted flat top pyramidal) geometry wells with 800x800 μ m (top surface) and 100 μ m depth were obtained. Optionally, the wells can be roughed or pitted. As described above, the chip edge was aligned against a raised surface on the stage to define the x and y coordinate systems with respect to the
15 capillary. (Alternatives include optical alignment, artificial intelligence pattern recognition routines, and dowel-pin based manual alignment). Into each well was dispensed 20 droplets (~5 nL) of 3-HPA matrix solution without analyte; for the 50% CH₃CN solution employed, evaporation times for each droplet were on the order of 5-10 seconds.
20 Upon solvent evaporation, each microdispensed matrix droplet as viewed under a 120X stereomicroscope generally appeared as an amorphous and 'milky' flat disk; such appearances are consistent with those of droplets from which the Figure 3b spectrum was obtained. Upon tip emptying, rinsing, and refilling with a 1.4 μ m aqueous solution of a 23-mer DNA
25 ($M_r(\text{calc}) = 6967$ Da), the capillary was directed above each of the 100 spots of matrix where 5 nL of the aqueous DNA solution was dispensed directly on top of the matrix droplets. Employing visualization via a CCD camera, it appeared that the aqueous analyte solution mixed with and re-dissolved the matrix (complete evaporation took ~ 10 sec at ambient

-72-

temperature and humidity). The amorphous matrix surfaces were converted to true micro-crystalline surfaces, with crystalline features on the order of $< 1 \mu\text{m}$.

Consistent with the improved crystallization afforded by the matrix re-dissolving method, mass spectrum acquisition appeared more reproducible than with pre-mixed matrix plus analyte solutions; each of the 100 five fmol spots of the 23-mer yielded interpreted mass spectra (Figure 11), with 99/100 parent ion signals having signal to noise ratios of > 5 ; such reproducibility was also obtained with the flat silicon and metallic surfaces tried (not shown). The Figure 11 spectra were obtained on a linear TOF instrument operated at 26 kV. Upon internal calibration of the top left spectrum (well 'k1') using the singly and doubly charged molecular ions, and application of this calibration file to all other 99 spectra as an external calibration (Figure 12), a standard deviation of < 9 Da from the average molecular weight was obtained, corresponding to a relative standard deviation of $\sim 0.1\%$.

Parallel dispensing with the robotic pintool

Arrays were made with offset printing as described above. The velocity of the X and Y stages are 35 inches/sec, and the velocity of the Z stage is 5.5 inches/sec. It is possible to move the X and Y stages at maximum velocity to decrease the cycle times, however the speed of the Z stage is to be decreased prior to surface contact with the wafer to avoid damaging it. At such axes speeds, the approximate cycle time to spot 16 elements (one tool impression of the same solutions) on all ten wafers is 20 seconds, so to make an array of 256 elements would take ~ 5.3 minutes. When placing different oligonucleotide solutions on the array, an additional washing step must be incorporated to clean the pin tip prior to dipping in another solution, thus the cycle time would increase to 25 seconds or 6.7 minutes to make 10 wafers.

-73-

Sample delivery by the tool was examined using radio-labeled solutions and the phosphorimager as described previously; it was determined that each pin delivers approximately 1 nL of liquid. The spot-to-spot reproducibility is high. An array of 256 oligonucleotide elements of varying sequence and concentration was made on flat silicon wafers using the pintool, and the wafer was analyzed by MALDI-TOF MS.

EXAMPLE 3

Use of High Density Nucleic Acid Immobilization to Generate Nucleic Acid Arrays

Employing the high density attachment procedure described in EXAMPLE 1, an array of DNA oligomers amenable to MALDI-TOF mass spectrometry analysis was created on a silicon wafer having a plurality of locations, e.g., depressions or patches, on its surface. To generate the array, a free thiol-containing oligonucleotide primer was immobilized only at the selected locations of the wafer (e.g., see Figure 13). Each location of the array contained one of three different oligomers. To demonstrate that the different immobilized oligomers could be separately detected and distinguished, three distinct oligonucleotides of differing lengths that are complementary to one of the three oligomers were hybridized to the array on the wafer and analyzed by MALDI-TOF mass spectrometry.

Oligodeoxynucleotides

Three sets of complementary oligodeoxynucleotide pairs were synthesized in which one member of the complementary oligonucleotide pair contains a 3'- or 5'-disulfide linkage (purchased from Operon Technologies or Oligos, Etc.). For example, Oligomer 1 (d(CTGATGCGTCGGATCATCTTTTT-SS); SEQ ID NO: 5) contains a 3'-disulfide linkage whereas Oligomer 2 (d(SS-CCTCTTGGGAAGTGTAGTATT); a 5'-disulfide derivative of SEQ ID

-74-

NO: 6) and Oligomer 3 (d(SS-GAATTCGAGCTCGGTACCCGG); a 5'-disulfide derivative of SEQ ID NO: 1) each contain a 5'-disulfide linkage.

The oligonucleotides complementary to Oligomers 1-3 were designed to be of different lengths that are easily resolvable from one another during MALDI-TOF MS analysis. For example, a 23-mer oligonucleotide (SEQ ID NO: 7) was synthesized complementary to a portion of Oligomer 1, a 12-mer oligonucleotide (SEQ ID NO: 8) was synthesized complementary to a portion of Oligomer 2 and a 21-mer (SEQ ID NO: 2; sequence denoted "MJM6" in EXAMPLE 1) was synthesized complementary to a portion of Oligomer 3. In addition, a fourth 29-mer oligonucleotide (SEQ ID NO: 9) was synthesized that lacks complementarity to any of the three oligomers. This fourth oligonucleotide was used as a negative control.

Silicon surface chemistry and DNA immobilization

(a) 4 x 4 (16-location) array

A 2 X 2 cm² silicon wafer having 256 individual depressions or wells in the form of a 16 X 16 well array was purchased from a commercial supplier (Accelerator Technology Corp., College Station, Texas). The wells were 800 X 800 μm^2 , 120 μm deep, on a 1.125 pitch. The silicon wafer was reacted with 3-aminopropyltriethoxysilane to produce a uniform layer of primary amines on the surface and then exposed to the heterobifunctional crosslinker SIAB resulting in iodoacetamido functionalities on the surface (e.g., see Figure 7).

To prepare the oligomers for coupling to the various locations of the silicon array, the disulfide bond of each oligomer was fully reduced using 10 mM TCEP as depicted in EXAMPLE 1, and the DNA resuspended at a final concentration of 10 μM in a solution of 100 mM phosphate buffer, pH 8.0. Immediately following disulfide bond reduction, the free-thiol group of the oligomer was coupled to the

-75-

iodoacetamido functionality at 16 locations on the wafer using the probe coupling conditions essentially as described in Figure 7. To accomplish the separate coupling at 16 distinct locations of the wafer, the entire surface of the wafer was not flushed with an oligonucleotide solution but, instead, an ~30-nl aliquot of a predetermined modified oligomer was added in parallel to each of 16 locations (i.e., depressions) of the 256 wells on the wafer to create a 4 x 4 array of immobilized DNA using a pin tool as described herein (see e.g., the Detailed Description and Example 4 provided herein).

Thus, as shown in Figure 13, one of modified Oligomers 1-3 was covalently immobilized to each of 16 separate wells of the 256 wells on the silicon wafer thereby creating a 4 x 4 array of immobilized DNA. For example, Oligomer 1 was conjugated at a well position in the upper left hand corner of the 4 x 4 array and Oligomer 2 was conjugated to the adjacent location, and so forth. An illustration of the completed array is shown in Figure 13.

In carrying out the hybridization reaction, the three complementary oligonucleotides and the negative control oligonucleotide were mixed at a final concentration of 10 μ M for each oligonucleotide in 1 ml of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) supplemented with 1 M NaCl, and the solution was heated at 65°C for 10 min. Immediately thereafter, the entire surface of the silicon wafer was flushed with 800 μ l of the heated oligonucleotide solution. The complementary oligonucleotides were annealed to the immobilized oligomers by incubating the silicon array at ambient temperature for 1 hr, followed by incubation at 4°C for at least 10 min. Alternatively, the oligonucleotide solution can be added to the wafer which is then heated and allowed to cool for hybridization. An illustration of the complementary oligonucleotides annealed to the

-76-

specific oligomers covalently immobilized at each location is shown in Figure 14.

The hybridized array was then washed with a solution of 50 mM ammonium citrate buffer for cation exchange to remove sodium and potassium ions on the DNA backbone (Pieles, U. et al., (1993) Nucl. Acids Res. 21:3191-3196). A 6-nl aliquot of a matrix solution of 3-hydroxypicolinic acid (0.7 M 3-hydroxypicolinic acid-10 % ammonium citrate in 50 % acetonitrile; see Wu et al., Rapid Commun. Mass Spectrom. 7:142-146 (1993)) was added to each location of the array using a piezoelectric pipette as described herein.

The solution was allowed to dry at ambient temperature and thereafter a 6-nl aliquot of water was added to each location using a piezoelectric pipette to resuspend the dried matrix-DNA complex, such that upon drying at ambient temperature the matrix-DNA complex forms a uniform crystalline surface on the bottom surface of each location.

MALDI-TOF MS analysis

The MALDI-TOF MS analysis was performed in series on each of the 16 locations of the hybridization array illustrated in Figure 14 essentially as described in EXAMPLE 1. The resulting mass spectrum of oligonucleotides that specifically hybridized to each of the 16 locations of the DNA hybridization array is shown in Figure 15. The mass spectrum revealed a specific signal at each location representative of observed experimental mass-to-charge ratio corresponding to the specific complementary nucleotide sequence.

For example, in the locations that have only Oligomer 1 conjugated thereto, the mass spectrum revealed a predominate signal with an observed experimental mass-to-charge ratio of 7072.4 approximately equal to that of the 23-mer; the theoretical mass-to-charge ratio of the 23-mer is 7072.6 Da. Similarly, specific hybridization of the 12-mer

-77-

oligonucleotide to the array, observed experimental mass-to-charge ratio of 3618.33 Da (theoretical 3622.4 Da), was detected only at those locations conjugated with Oligomer 2 whereas specific hybridization of MJM6 (observed experimental mass-to-charge ratio of 6415.4) was
5 detected only at those locations of the array conjugated with Oligomer 3 (theoretical 6407.2 Da).

None of the locations of the array revealed a signal that corresponds to the negative control 29-mer oligonucleotide (theoretical mass-to-charge ratio of 8974.8) indicating that specific target DNA
10 molecules can be hybridized to oligomers covalently immobilized to specific locations on the surface of the silicon array and a plurality of hybridization assays may be individually monitored using MALDI-TOF MS analysis.

(b) 8 x 8 (64-location) array

15 A 2 X 2 cm² silicon wafer having 256 individual depressions or wells that form a 16 X 16 array of wells was purchased from a commercial supplier (Accelerator Technology Corp., College Station, Texas). The wells were 800 X 800 μm^2 , 120 μm deep, on a 1.125 pitch. The silicon wafer was reacted with 3-aminopropyltriethoxysilane
20 to produce a uniform layer of primary amines on the surface and then exposed to the heterobifunctional crosslinker SIAB resulting in iodoacetamido functionalities on the surface (e.g., see Figure 7).

Following the procedures described above for the preparation of the 16-location DNA array, Oligomers 1-3 were immobilized to 64
25 locations forming an 8 X 8 array on the 256 well silicon wafer, hybridized to complementary oligonucleotides and analyzed by MALDI-TOF MS analysis. Figure 16 shows the mass spectrum of the 64-location DNA array analyzed in series by MALDI-TOF analysis. As shown for the 16-location array, specific hybridization of the complementary

-79-

SP6 RNA Polymerase Transcription

In vitro transcription of the nicked DNA templates was performed in 20 μ l reactions of 40 mM Tris-HCl (pH 7.0), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 1 unit/ μ l RNasin (Promega), 5 mM rNTP, 5 μ Ci [α - 32 P]rCTP, 1 unit/ μ l SP6 RNA polymerase (Amersham, Arlington Heights, IL) at 37°C for 30 minutes. Abortive and full length RNA transcripts were separated by gel electrophoresis and quantified by measuring the radioactivity of individual RNA fragments by drying the polyacrylamide gel and measuring the radioactivity as compared to a known standard using a PhosphorImager (Molecular Dynamics, Inc.). The efficiency of full-length RNA transcription of a nicked DNA template was calculated as a percentage the moles of full length RNA transcribed from a DNA template containing no nicks. The nick by pass efficiency of a nicked DNA template was calculated as a percentage of the moles full-length RNA transcript and the moles of RNA transcript stalled at the nick.

As illustrated in Table 1, the transcription of a full-length RNA proceeds with 88-94% efficiency when a nick is introduced into the coding strand after nucleotides +7, +8, +9 or +19 relative to the start of transcription.

TABLE 1

| | 6-mer RNA (%) | Nick-stalled RNA (%) | Full-Length RNA (%) | Nick-Bypass Efficiency (%) |
|--------|------------------|-------------------------|------------------------|----------------------------------|
| Ref. | 320 \pm 93 | - | (100) | 100 \pm 0 |
| 1N+7U | 193.2 \pm 46.3 | 12.45.4 | 95.8 \pm 6.5 | 88.5 \pm 5.1 |
| 1N+8U | 187.4 \pm 39.9 | 5.3 \pm 4.2 | 87.1 \pm 4.3 | 94.3 \pm 0.5 |
| 1N+9U | 232.7 \pm 45.6 | 10.1 \pm 1.1 | 65.7 \pm 4.4 | 87.6 \pm 1.8 |
| 1N+19U | 278.6 \pm 33.8 | 6.6 \pm 0.3 | 64.8 \pm 4.5 | 90.9 \pm 0.4 |

-80-

EXAMPLE 5**DNA Sequencing using T7 RNA Polymerase****5 1. Design of Template and Primer Sequences**

All primers were synthesized on a commercially available DNA Synthesizer using conventional phosphoroamidite chemistry (Sinha et al. (1984) Nucleic Acid Res. 12:4539). In vitro RNA transcription was performed on a synthetic 276 nucleotide double-stranded DNA template
10 (SEQ ID No. 13, which includes a T7 promoter).

2. DNA Sequencing

DNA sequencing of a target DNA template was carried out in 10 μ l reactions of 40 mM Tris-HCl (pH 7.9), 6 mM $MgCl_2$, 2 mM neutralized spermidine, 5 mM dithiothreitol, 300 nM rCTP, 300 nM rATP, 300 nM
15 rUTP, 300 nM rITP, 600 nM rGMP, 10 μ Ci (α - ^{32}P) rCTP or UTP, 10-300, generally 10-50 μ M 3'-deoxynucleotide, 0.5-1.0 pmole linearized or supercoiled DNA template, 4 units RNasin (Promega), 10 unit/ μ l T7 RNA polymerase (USB) at 37°C for 30 minutes. Chain termination fragments RNA transcripts were separated by gel electrophoresis using 8% or 13%
20 polyacrylamide gel.

DNA sequencing ladders of RNA terminated fragment were generated up to 180-200 bases. This example shows that T7 RNA polymerase can specifically incorporate modified 3'-deoxyribonucleosides triphosphates as a base-specific chain terminator to generate nested RNA
25 transcripts for sequencing nucleic acids.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-81-

WHAT IS CLAIMED:

1. A method for determining the sequence of a target nucleic acid molecule, comprising:
 - a) providing a nucleic acid molecule comprising a promoter and
5 target nucleic acid sequence operatively linked to the promoter, wherein the nucleic acid molecule is immobilized on a solid support;
 - b) transcribing the promoter-containing nucleic acid molecule with an RNA polymerase that recognizes the promoter under conditions whereby a nested set of RNA transcripts from the target is produced;
10 and
 - c) determining the molecular weight values of the transcripts by mass spectrometry and thereby determining the nucleic acid sequence of the target molecule.
2. The method of claim 1, wherein the promoter-containing
15 nucleic acid molecule is immobilized via the 5' end of the coding strand or the 3' end of the non-coding strand.
3. The method of claim 1, wherein the nucleic acid is DNA or PNA.
4. The method of any of claims 1-3, wherein the nucleic acid molecule comprising a promoter and target is prepared by:
20 hybridizing a nucleic acid promoter-containing probe having a coding and noncoding strand to a nucleic acid molecule comprising a target nucleic acid sequence, wherein hybridization occurs between a single stranded region of the promoter-containing probe and a complementary single stranded region at one end of the molecule
25 comprising the target.
5. The method of claim 4, wherein the molecule comprising the target is single-stranded and the portion comprising the complementary single-stranded region is at the 3' end of the target-containing molecule;

-82-

and the promoter containing probe molecule is double-stranded with a single-stranded region of at least 5 nucleotides at the 3' end of the coding strand, wherein at least 5 contiguous nucleotides are complementary to the single-stranded portion of the target-containing molecule.

6. The method of claim 4, wherein:
the molecule comprising the target is double-stranded except for a single-stranded region at one end that comprises at least 5 contiguous nucleotides complementary to the single-stranded portion of the promoter-containing molecule; and
the single-stranded portion of the promoter-containing molecule is at same end, relative to the promoter, that the single-stranded region is on the target-containing molecule.
7. The method of any of claims 4-6, wherein the promoter-containing molecule is immobilized on the solid support prior to hybridization of the target-containing molecule.
8. The method of any of claims 4-6, wherein hybridization is performed prior to immobilization of the promoter-containing molecule on the solid support.
9. The method of any of claims 4-8, wherein nicks in the resulting molecules are ligated prior to transcription by the RNA polymerase.
10. The method of any of claims 1-9, wherein the nucleic acid molecule comprising a promoter and target nucleic acid is transcribed with an RNA polymerase to produce a plurality of base-specifically terminated RNA transcripts.
11. The method of any of claims 1-9, wherein the nucleic acid molecule comprising a promoter and target nucleic acid is transcribed with an RNA polymerase to produce a set of RNA transcripts that are

-83-

then specifically cleaved with base-specific ribonucleases to produce a set of cleaved fragments whose molecule weight values are assessed by mass spectrometry.

12. The method of any of claims 1-11 wherein the sequence of
5 all or a portion of the target nucleic acid is determined by the RNA transcripts according to molecular weight.

13. The method of any of claims 1-12, wherein the nucleic acid molecule comprising the promoter is produced by immobilizing a single-stranded molecule that comprises a promoter or the complement of a
10 promoter and hybridizing a fragment that comprises the complement thereof, whereby the resulting molecule comprises a double-stranded region containing a promoter and a single-stranded portion of at least 5 nucleotides.

14. The method of any of claims 1-13, wherein the promoter is
15 selected from the group consisting of phage, eukaryotic and prokaryotic promoters.

15. The method of claim 14, wherein the promoter is selected from archaeobacteria, eubacteria, bacteriophages, DNA viruses, RNA viruses, plants, plant viruses and animal promoters.

20 16. The method of any of claims 1-15, wherein the RNA polymerase is a DNA-dependent RNA polymerase or an RNA-dependent RNA polymerase.

17. The method of any of claims 1-16, wherein the RNA polymerase is selected from the group consisting of phage, eukaryotic
25 and prokaryotic polymerases.

18. The method of claim 17, wherein the polymerases is selected from archaeobacteria, eubacteria, bacteriophages, DNA viruses, RNA viruses, plants, plant viruses and animal polymerases.

-84-

19. The method of any of claims 1-18, wherein the RNA polymerase is selected from the group consisting of Escherichia coli RNA polymerase, T7 RNA polymerase, SP6 RNA polymerase, T3 RNA polymerase and Q β replicase.
- 5 20. The method of any of claims 1-19, wherein prior to immobilization of the nucleic acid, the surface of the support is derivatized by reacting the surface with an aminosilane to produce primary amines on the surface of the support.
21. The method of claim 20, wherein the aminosilane is 3-
10 aminopropyltriethoxysilane.
22. The method of claim 21, further comprising reacting the primary amines on the surface of the support with a thiol-reactive cross-linking reagent to form a thiol-reactive solid support.
23. The method of claim 22, wherein the thiol-reactive cross-
15 linking reagent is N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB).
24. The method of claim 22, wherein the immobilization of the nucleic acid probe to a solid support is effected by reacting the thiol-reactive solid support with a nucleic acid probe having a free 5'- or 3'-
20 thiol group, whereby a covalent bond between the thiol group and the thiol-reactive solid support is formed.
25. The method of any of claims 1-21, wherein the nucleic acid molecule is immobilized to the surface the solid support at a density of at least 20 fmol/mm² by a covalent linkage.
26. The method of any of claims 1-25, wherein the nucleic acids
25 are immobilized on the surface of the solid support in the form of an array.
27. The method of any of claims 1-26, wherein the solid support is silicon or silicon coated.

-85-

28. The method of any of claims 1-27, wherein the surface comprises a plurality of wells comprising the immobilized nucleic acid molecule.

29. The method of claim 28, wherein the wells have a rough
5 interior surface.

30. The method of claim 28, wherein the solid support has a rough surface.

31. The method of claim 28, wherein the surface of the wells is pitted.

10 32. The method of any of claims 1-31 wherein the transcription reactions are performed in the presence of a ribonucleoside triphosphate analog, whereby the resulting RNA transcripts have reduced secondary structure or the fidelity of termination and turnover of the RNA polymerase enzyme is increased compared to
15 RNA transcripts formed from ribonucleoside triphosphatase.

33. The method of claim 32, wherein the transcription reactions include one or more analogs selected from among inosine 5'-triphosphate (ITP), 4-thio uridine 5'-triphosphate (UTP), 5-bromo UTP and 5'-iodo CTP.

20 34. The method of any of claims 1-33, wherein the nucleic acid is immobilized to the solid support via a linker.

35. The method of claim 34, wherein the linkage formed by the linker is reversible.

36. The method of claim 34, wherein the linkage formed by the linker is irreversible.

25 37. The method of claim 34, wherein the linkage formed by the linker is photolabile, acid labile or chemically cleavable.

38. The method of any of claims 1-37, wherein transcription is performed in the presence of one or more 3'-deoxyribonucleotides.

-86-

39. The method of any of claims 1-38, further comprising:
adding a matrix material for mass spectrometry to the surface of
the support before or after immobilization of the nucleic acid.
40. The method of any of claims 1-39, wherein the RNA
5 transcripts are conditioned prior to performing mass spectrometry.
41. The method of any of claims 1-4, wherein the hybridization
of the nucleic acid containing the target sequence to the promoter-
containing molecule results in the formation of a nick in the coding strand
corresponding to positions beyond +8 relative to the start of
10 transcription from the promoter.
42. The method of claim 24, wherein the nick is at position
+7, +8, +9 or +19.
43. The method of claim 1, comprising:
- a) immobilizing a nucleic acid promoter-containing probe on a
15 solid support, wherein the nucleic acid promoter-containing probe
comprises at least 5 nucleotides at the 3'-end of the coding strand that is
complementary to a single-stranded region at the 3'-end of the target
nucleic acid;
- b) hybridizing the nucleic acid to be sequenced to the
20 immobilized nucleic acid probe;
- c) transcribing the target nucleic acid with an RNA polymerase
to produce a plurality of base-specifically terminated RNA transcripts,
wherein the RNA polymerase recognizes the promoter; and
- d) determining the molecular weight value of each
25 base-specifically terminated RNA transcript by mass spectrometry.
44. A method of identifying transcriptional terminator sequences
or attenuator sequences, comprising:
- a) immobilizing a nucleic acid promoter-containing probe on a
solid support, wherein the nucleic acid promoter-containing probe

-87-

comprises at least 5 nucleotides at the 3'-end of the coding strand that is complementary to a single stranded region at the 3'-end of the target nucleic acid;

- b) hybridizing the nucleic acid to be sequenced to the
- 5 immobilized nucleic acid probe;
- c) transcribing the target nucleic acid with an RNA polymerase to produce a sequence-terminated RNA transcript, wherein the RNA polymerase recognizes the promoter; and
- d) determining the molecular weight value of the RNA
- 10 transcript by mass spectrometry, whereby the terminator sequence or attenuator is identified.

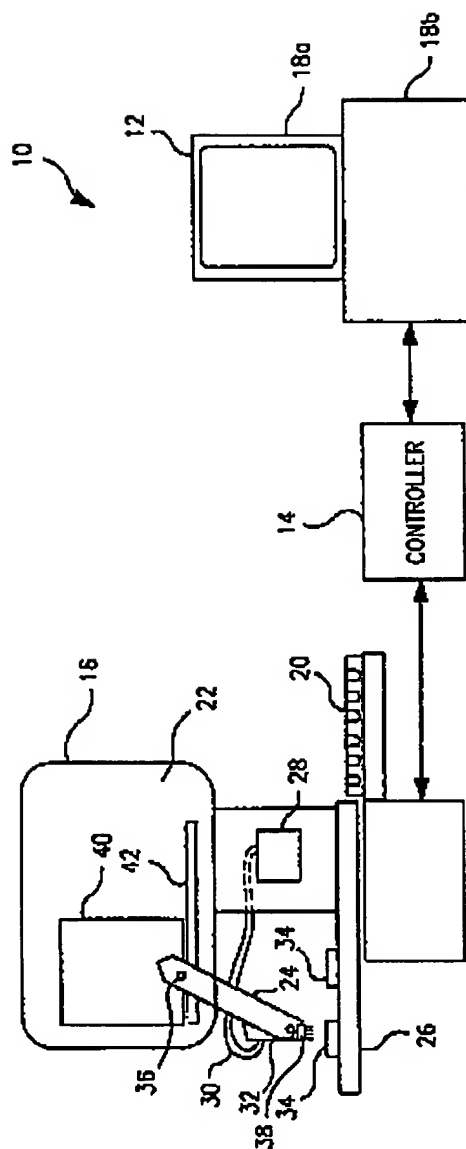


FIG. 1

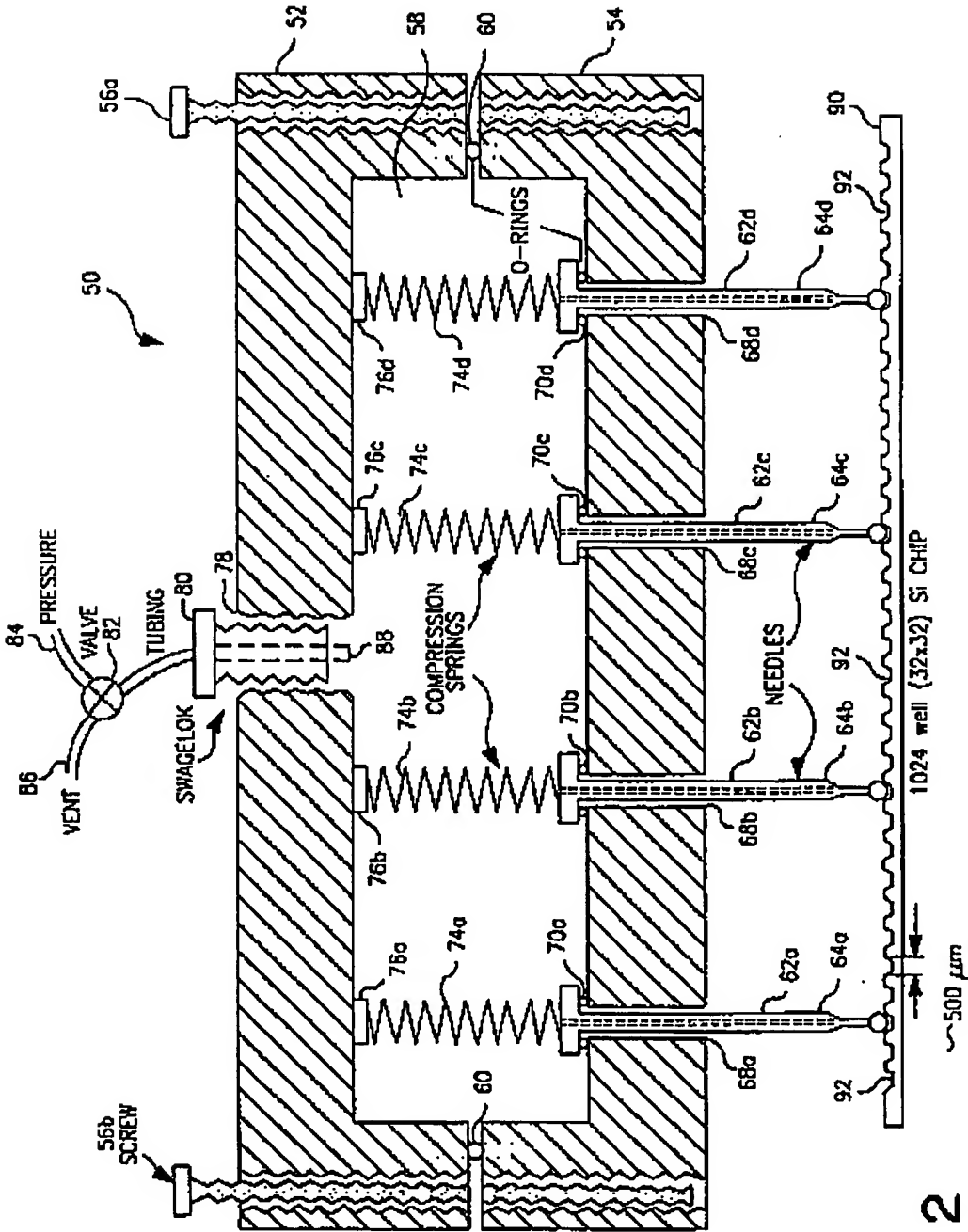


FIG. 2

3/17

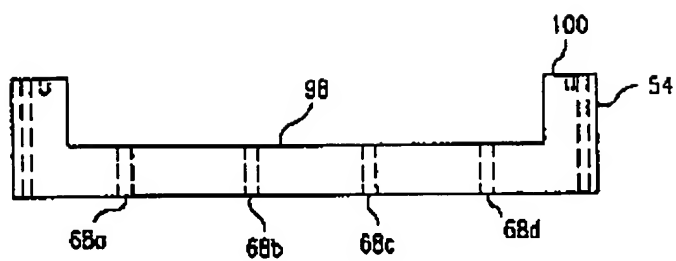


FIG. 3

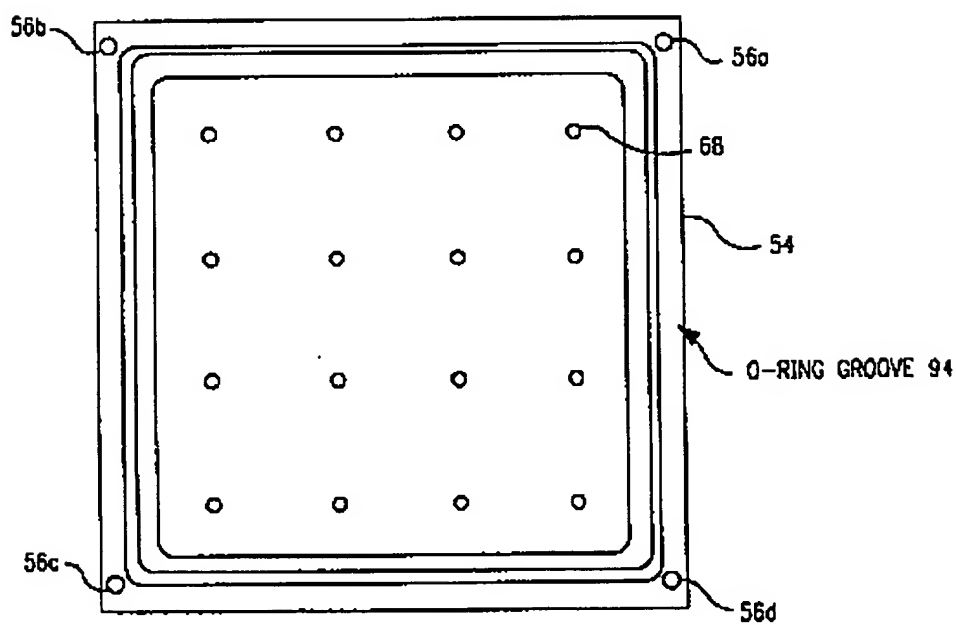


FIG. 4

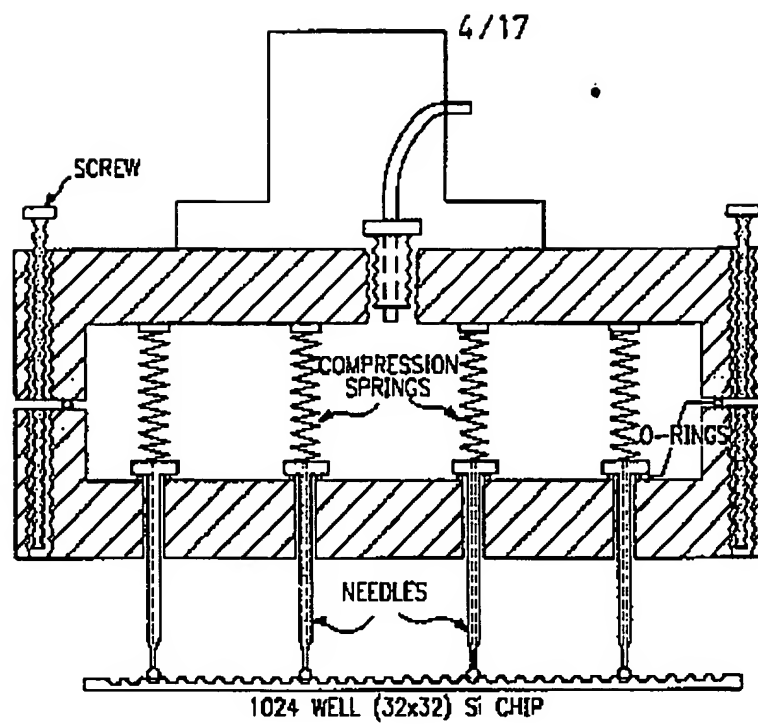


FIG. 5A

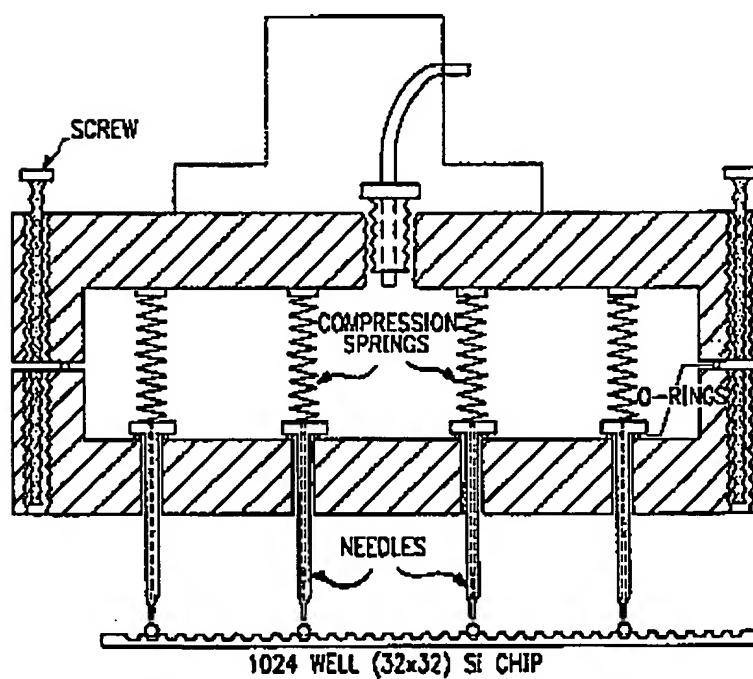


FIG. 5B

5/17

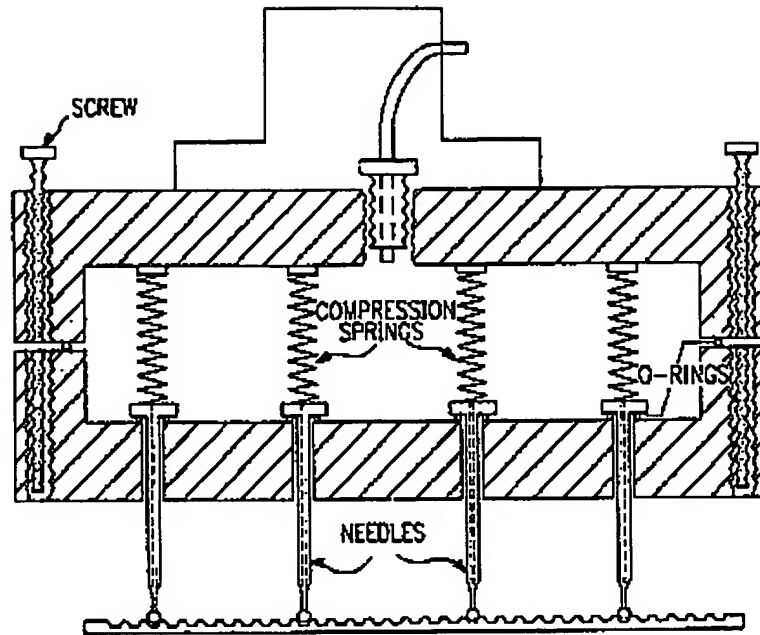


FIG. 5C

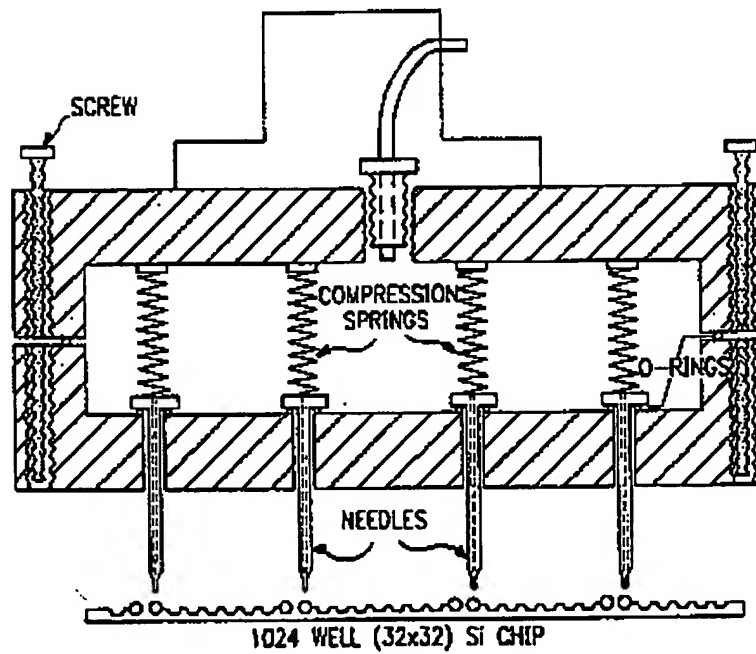


FIG. 5D

6/17

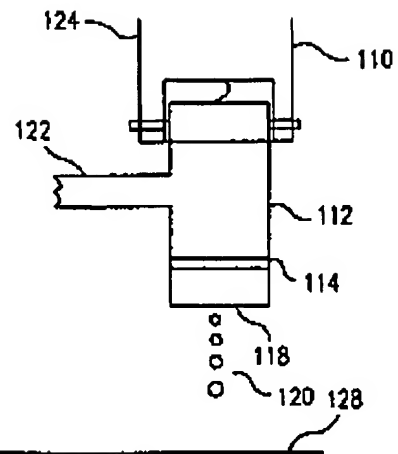


FIG. 6A

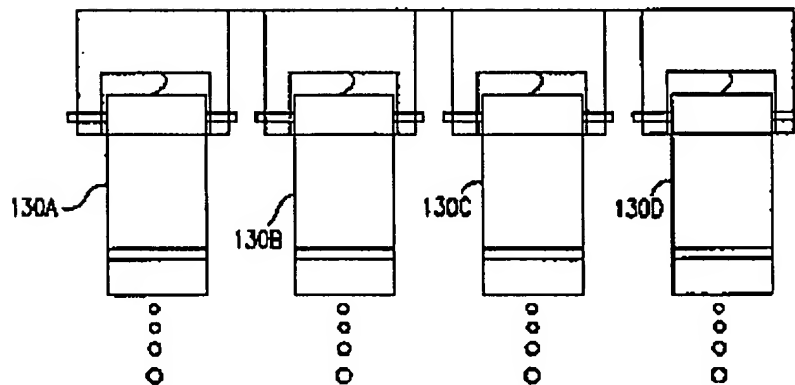


FIG. 6B

7/17

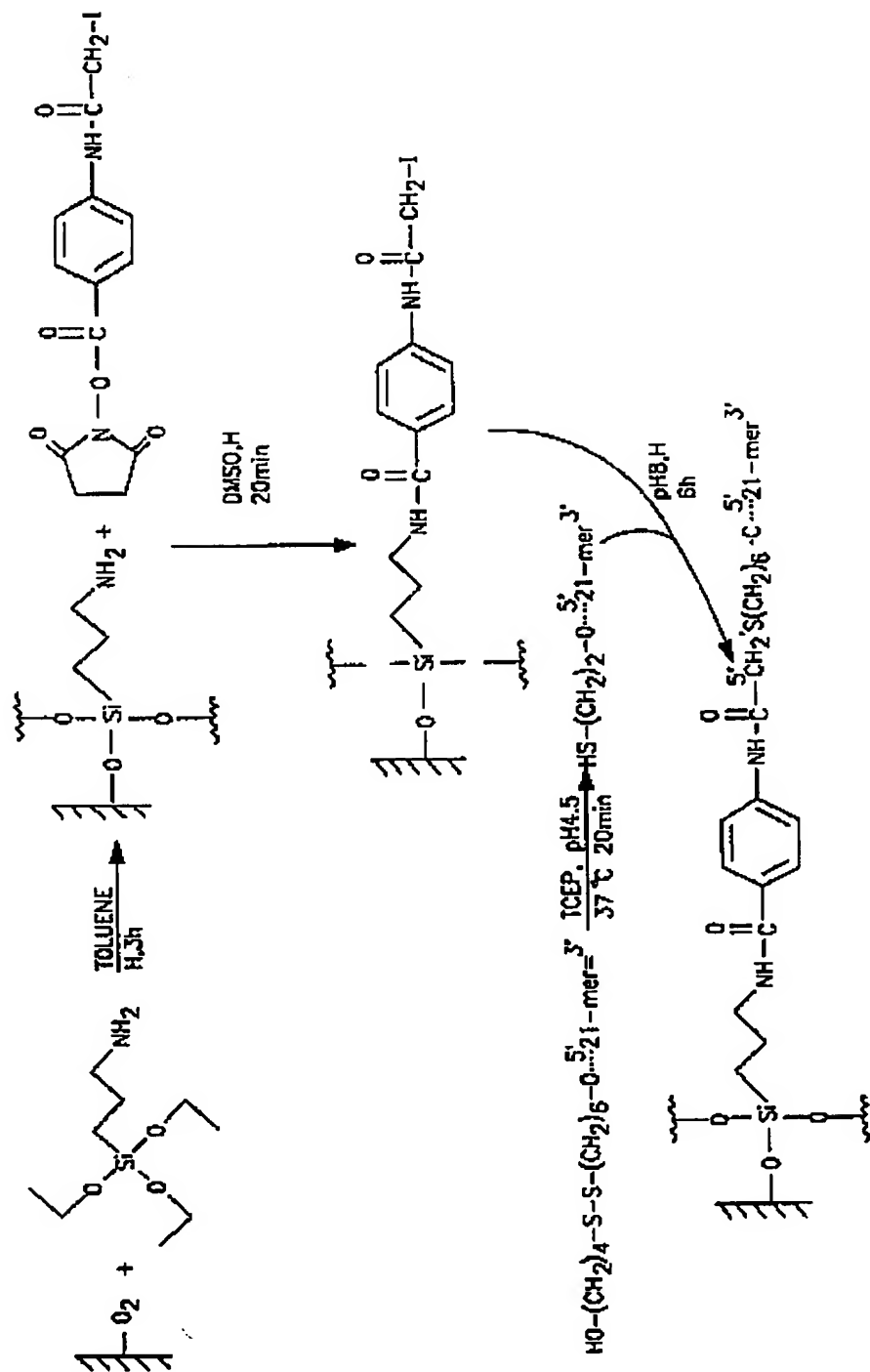


Fig. 7

8/17

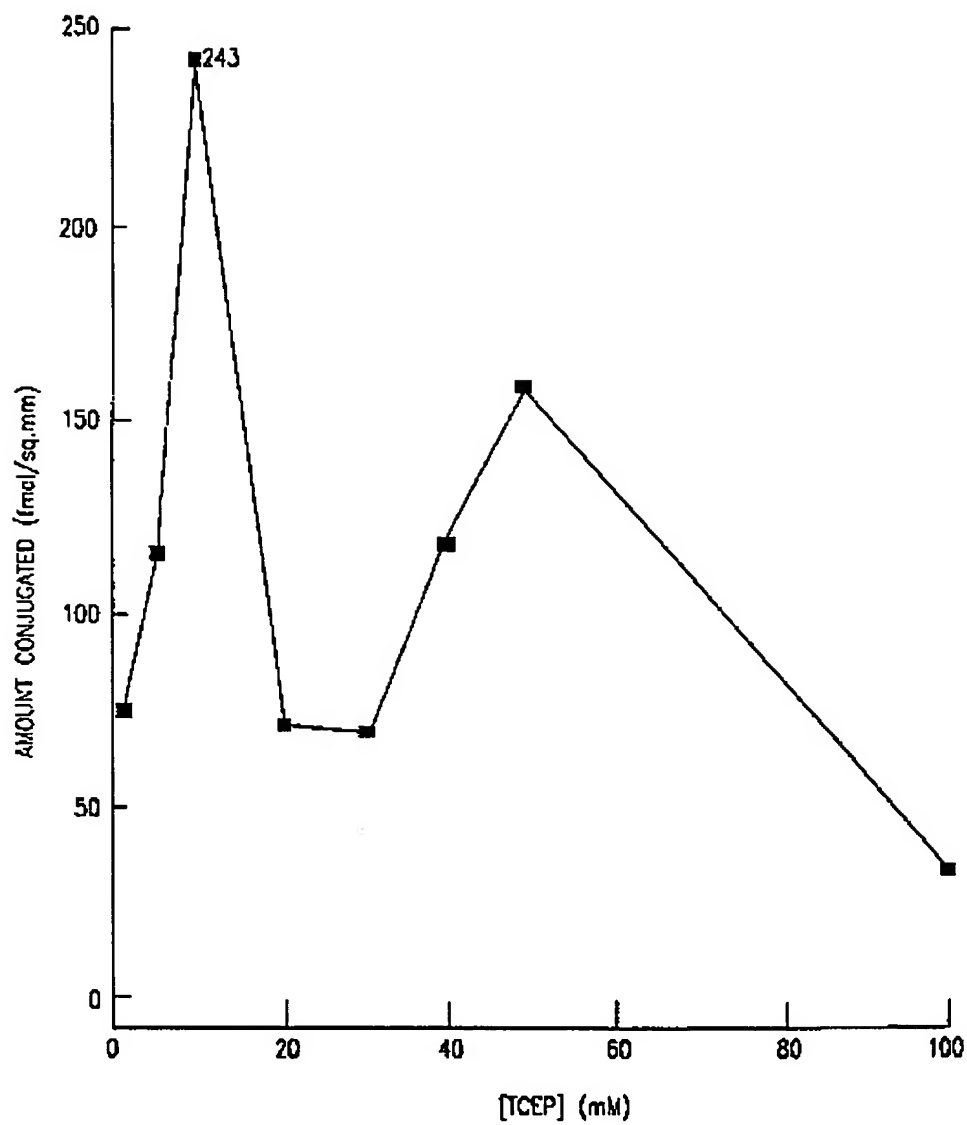


FIG. 8

9/17

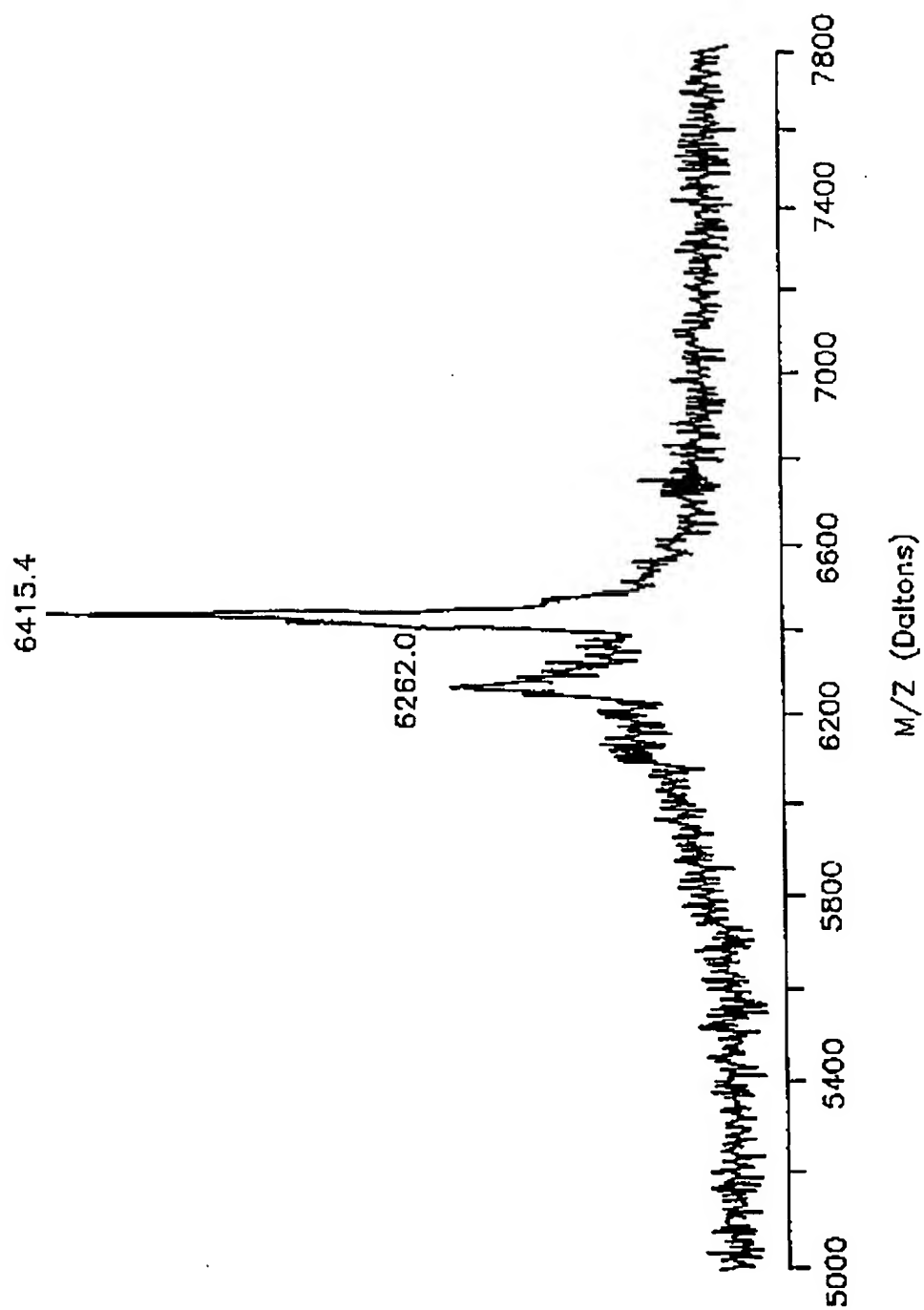


FIG. 9

10/17

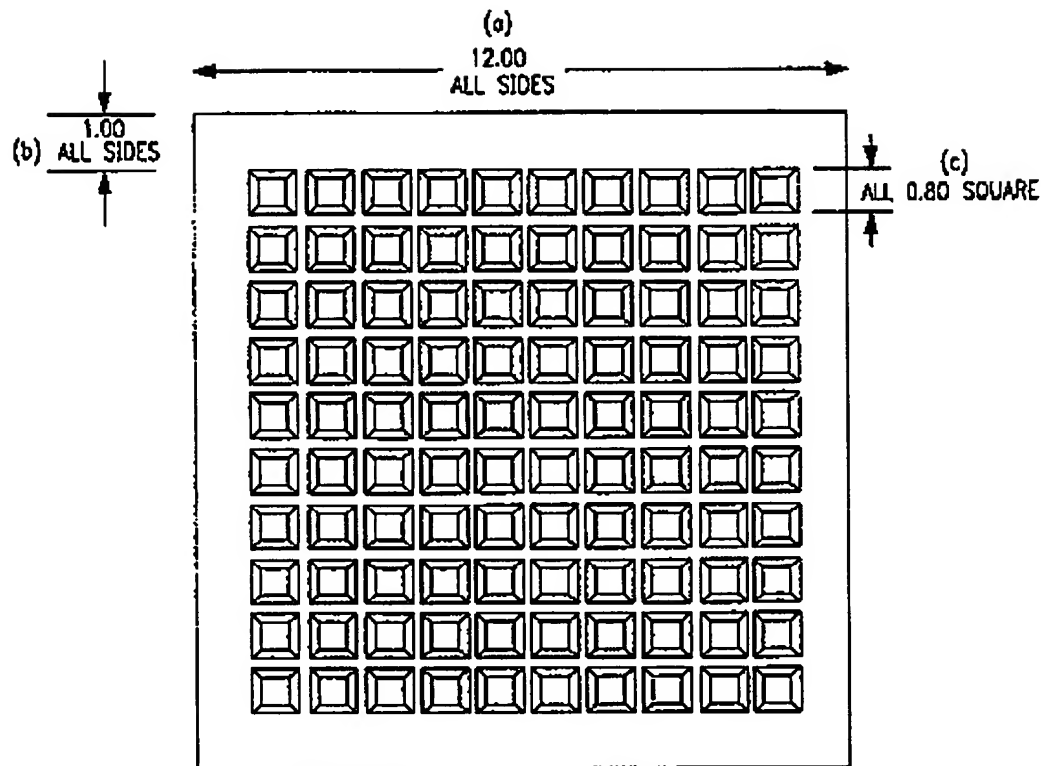


FIG. 10

11/17

23-MER (6nL of 1.4uM = 8.6fmol)
10x10 850x850um (89um DEPTH) WELLS.

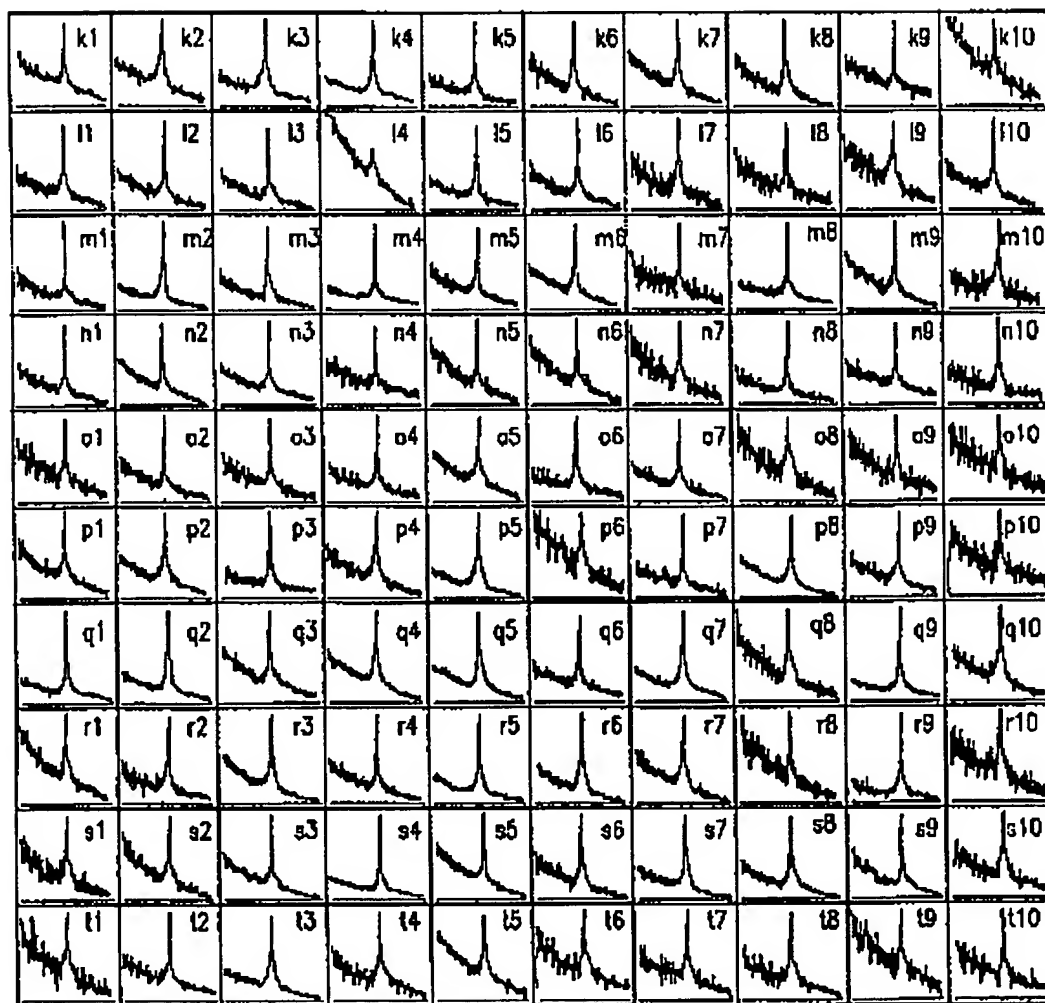


FIG. 11

12/17

| | | | | | | | | | |
|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| k1 6968 Da 170 RP | k2 6968 Da 100 RP | k3 6988 Da 90 RP | k4 6977 Da 100 RP | k5 6971 Da 170 RP | k6 6968 Da 110 RP | k7 6972 Da 160 RP | k8 6978 Da 110 RP | k9 6952 Da 250 RP | k10 6965 Da 300 RP |
| l1 6965 Da 130 RP | l2 6989 Da 140 RP | l3 6982 Da 210 RP | l4 6996 Da 50 RP | l5 6982 Da 160 RP | l6 6968 Da 180 RP | l7 6984 Da 130 RP | l8 6968 Da 200 RP | l9 6996 Da 80 RP | l10 6968 Da 100 RP |
| m1 6966 Da 190 RP | m2 6979 Da 120 RP | m3 6975 Da 120 RP | m4 6968 Da 190 RP | m5 6976 Da 110 RP | m6 6986 Da 120 RP | m7 6973 Da 160 RP | m8 6978 Da 160 RP | m9 6975 Da 230 RP | m10 6955 Da 250 RP |
| n1 6961 Da 340 RP | n2 6971 Da 180 RP | n3 6970 Da 150 RP | n4 6960 Da 300 RP | n5 6985 Da 120 RP | n6 6953 Da 210 RP | n7 6971 Da 140 RP | n8 6962 Da 160 RP | n9 6957 Da 150 RP | n10 6960 Da 160 RP |
| o1 6965 Da 140 RP | o2 6960 Da 230 RP | o3 6976 Da 200 RP | o4 6953 Da 250 RP | o5 6983 Da 110 RP | o6 6967 Da 250 RP | o7 6970 Da 150 RP | o8 6973 Da 70 RP | o9 6953 Da 140 RP | o10 6952 Da 140 RP |
| p1 6976 Da 140 RP | p2 6981 Da 90 RP | p3 6972 Da 180 RP | p4 6969 Da 90 RP | p5 6984 Da 130 RP | p6 6968 Da 100 RP | p7 6958 Da 290 RP | p8 6981 Da 100 RP | p9 6978 Da 110 RP | p10 6965 Da 150 RP |
| q1 6976 Da 170 RP | q2 6985 Da 100 RP | q3 6990 Da 120 RP | q4 6989 Da 90 RP | q5 6984 Da 90 RP | q6 6969 Da 170 RP | q7 6979 Da 70 RP | q8 6968 Da 140 RP | q9 6973 Da 120 RP | q10 6950 Da 120 RP |
| r1 6966 Da 130 RP | r2 6960 Da 150 RP | r3 6969 Da 100 RP | r4 6964 Da 180 RP | r5 6965 Da 130 RP | r6 6970 Da 110 RP | r7 6972 Da 90 RP | r8 6939 Da 130 RP | r9 6951 Da 230 RP | r10 6965 Da 200 RP |
| s1 6963 Da 130 RP | s2 6953 Da 210 RP | s3 6970 Da 120 RP | s4 6971 Da 170 RP | s5 6957 Da 130 RP | s6 6956 Da 160 RP | s7 6966 Da 140 RP | s8 6975 Da 120 RP | s9 6951 Da 230 RP | s10 6969 Da 120 RP |
| t1 6974 Da 90 RP | t2 6958 Da 160 RP | t3 6959 Da 120 RP | t4 6952 Da 100 RP | t5 6959 Da 110 RP | t6 6954 Da 100 RP | t7 6950 Da 160 RP | t8 6974 Da 140 RP | t9 6967 Da 150 RP | t10 6950 Da 230 RP |

LASER POWER = 41000 FOR ALL SPECTRA.
EACH SPECTRUM THE SUM OF 10-30 SINGLE SHOTS.

FIG. 12

13/17

| | | | |
|--|---|---|---|
| <p>~slab~³TTTTCTTACTAGAGCTGCGTACGTC⁵</p> <p>Oligomer 1</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> | <p>~slab~³CCCTCTTGAGAACTGTGTATAT⁵</p> <p>Oligomer 2</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> |
| <p>~slab~³TTTTCTTACTAGAGCTGCGTACGTC⁵</p> <p>Oligomer 1</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> | <p>~slab~³CCCTCTTGAGAACTGTGTATAT⁵</p> <p>Oligomer 2</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> |
| <p>~slab~³TTTTCTTACTAGAGCTGCGTACGTC⁵</p> <p>Oligomer 1</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> | <p>~slab~³CCCTCTTGAGAACTGTGTATAT⁵</p> <p>Oligomer 2</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> |
| <p>~slab~³TTTTCTTACTAGAGCTGCGTACGTC⁵</p> <p>Oligomer 1</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> | <p>~slab~³CCCTCTTGAGAACTGTGTATAT⁵</p> <p>Oligomer 2</p> | <p>~slab~³GAATTCGAGCTCGGTACGTC⁵</p> <p>Oligomer 3</p> |

FIG. 13

14/17

| | | | |
|--|--|--|--|
| <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 1</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 2</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 3</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 1</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> |
| <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 3</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 1</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 2</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 3</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> |
| <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 2</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 3</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 1</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 2</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> |
| <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 1</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 2</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 3</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> | <p>Sub-3' TTTTCTTCTAGAGCTGAGAGTGC 21mer</p> <p>Oligomer 1</p> <p>Sub-3' GATGATCCGACGACGACGACGAGATG 21mer</p> |

FIG. 14

15/17

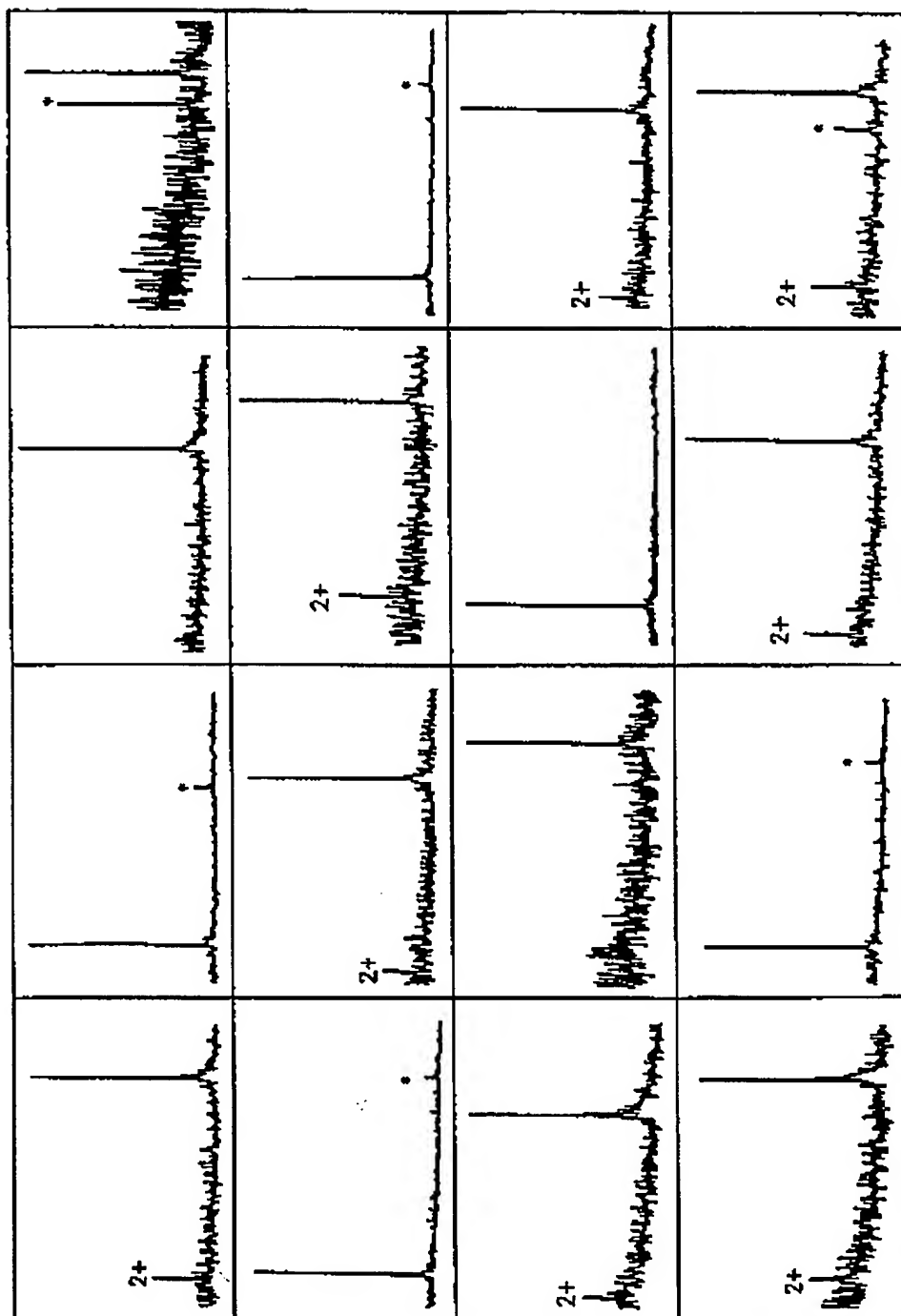


FIG. 15

16/17

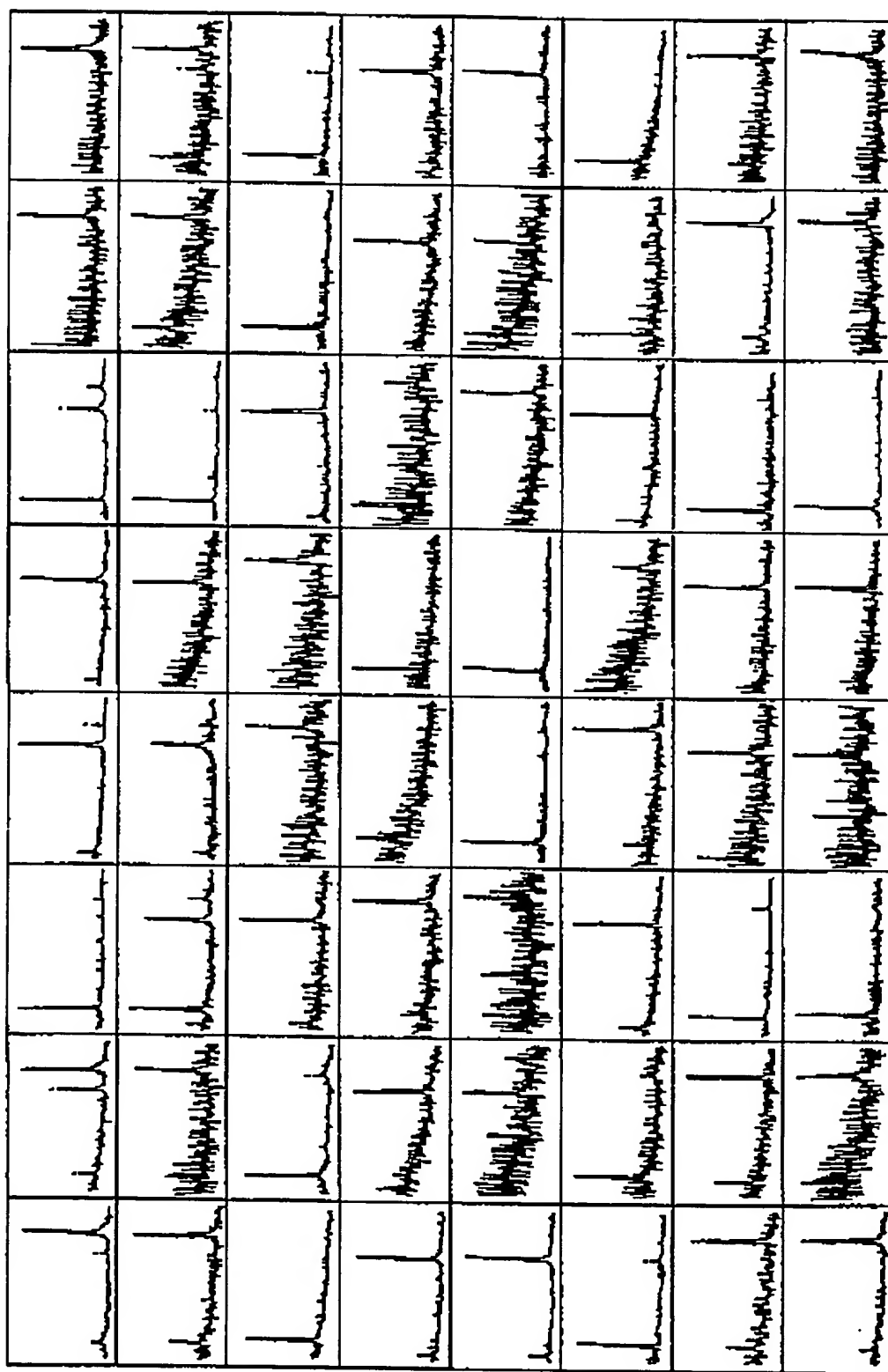


FIG. 16

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT:

(A) NAME: SEQUENOM, INC.
(B) STREET: 11555 Sorrento Valley Road
(C) CITY: San Diego
(D) STATE: California
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 92121

(i) APPLICANT:

(A) NAME: KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
(B) STREET: 373-1 Yusong-dong
(C) CITY: Taejeon 305 701
(D) STATE: Yusong-gu
(E) COUNTRY: Republic of Korea
(F) POSTAL CODE (ZIP): 305701

(i) INVENTOR

(A) NAME: Changwon Kang
(B) STREET: Expo Apt. 305-401, 161-1 Jeonmin-dong
(C) CITY: Taejeon 305-390
(D) STATE: Yusong-gu
(E) COUNTRY: REPUBLIC OF KOREA
(F) POSTAL CODE (ZIP):

(i) INVENTOR

(A) NAME: Young-Soo Kwon
(B) STREET: 25-2 Mangwol-dong
(C) CITY: Kwangju 500-500
(D) STATE: Puk-gu
(E) COUNTRY: Republic of Korea
(F) POSTAL CODE (ZIP):

(i) INVENTOR:

(A) NAME: Young Tae Kim
(B) STREET: Sungwon Apt. 2 204, 162 Songpa-2-dong
(C) CITY: Seoul 138-172
(D) STATE: Songpa-gu
(E) COUNTRY: REPUBLIC
(F) POSTAL CODE (ZIP):

(i) INVENTOR:

(A) NAME: Hubert Köster
(B) STREET: 836 Via Mallorca Drive
(C) CITY: La Jolla
(D) STATE: California
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 92037

(i) INVENTOR:

(A) NAME: Daniel P. Little
(B) STREET: 393 Glendale Lake Rd.
(C) CITY: Patton
(D) STATE: Pennsylvania
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 18668

(1) INVENTOR:

(A) NAME: Maryanne O'Donnell
(B) STREET:
(C) CITY:
(D) STATE:
(E) COUNTRY:
(F) POSTAL CODE (ZIP):

(2) INVENTOR:

(A) NAME: Goubing Xiang
(B) STREET: 11391 Zapata Ave.
(C) CITY: San Diego
(D) STATE: California
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 92126

(3) INVENTOR:

(A) NAME: Charles Cantor
(B) STREET:
(C) CITY:
(D) STATE: California
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP):

(4) INVENTOR:

(A) NAME: David Lough
(B) STREET:
(C) CITY:
(D) STATE:
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP):

(ii) TITLE OF INVENTION: MASS SPECTROMETRIC METHODS FOR SEQUENCING
NUCLEIC ACIDS

(iii) NUMBER OF SEQUENCES: 13

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Heller Ehrman White & McAuliffe
(B) STREET: 4250 Executive Square, 7th Floor
(C) CITY: La Jolla
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92037

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: ASCII

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/990,651
(B) FILING DATE: 15-DEC-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Seidman, Stephanie L
(B) REGISTRATION NUMBER: 33,779
(C) REFERENCE/DOCKET NUMBER: 24736-2007

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 619-450-8400
(B) TELEFAX: 619-450-8499

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCGAGC TCGGTACCCG G

21

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCGGGTACCG AGCTCGAATT C

21

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCGATCGAT CG

12

(2) INFORMATION FOR SEQ ID NO:4:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATCGATCGAT CGATCCATCC ATCGATCG

28

(2) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTGATGCGTC GGATCACTTT TTTT

24

(2) INFORMATION FOR SEQ ID NO:6:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCTCTTGGGA ACTGTGTAGT ATT

23

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATGATCCGA CGCATCAGAA TGT

23

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AATACTACAC AG

12

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATCTAGCTG GGCCGAGCTA GGCCCTTGA

29

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ATAGACGCTG CTGGACGGCA CCTTCTCC AACTTCTAT ACTCTACCT AAATC 55

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GATTTAGGTG ACACATAGA AGTCT

25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TGGAGAAGGG TGCCCTCCAG CAGCGTCTAT

30

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 276 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GCTCTAATAC GACTCAGTAT AGGGAGACAA GCTTGCATGC CTGCAGSTCG ACTCTAGAGG 60
ATCCCGGGGT ACUGAGCTCG AATTCCTGGC AGTTTATGGC GGGCGTCTGC CACCTTCCGG 120
GCCGTTGCTT CGCAACGTC AAATCCCGCT CCCGCCGATT TCTCCTACTC AGGAGAGCGT 180
TCACCCGACA AACACAGAT AAAAAAAG CCGAGTCTTT CGACGGGCC TTTGGTTTAA 240
TTTGATGCCG GGAATTCGT ATTCTATTCT ATAGTG 276
```

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PC7/US 98/26718

| Patent document cited in search report | | Publication date | Patent family member(s) | Publication date |
|---|---|---------------------|---|--|
| WO 9632504 | A | 17-10-1996 | AU 5544696 A CA 2218188 A EP 0830460 A JP 11503611 T | 30-10-1996 17-10-1996 25-03-1998 30-03-1999 |
| WO 9737041 | A | 09-10-1997 | AU 2217597 A | 22-10-1997 |
| EP 0785278 | A | 23-07-1997 | CA 2204378 A WO 9614434 A | 17-05-1996 17-05-1996 |
| US 5571669 | A | 05-11-1996 | NONE | |
| WO 8907149 | A | 10-08-1989 | EP 0358737 A | 21-03-1990 |
| WO 9605323 | A | 22-02-1996 | FR 2723591 A AT 165870 T DE 69502402 D DE 69502402 T EP 0775217 A | 16-02-1996 15-05-1998 10-06-1998 10-12-1998 28-05-1997 |

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 98/26718

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols):
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used):

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| Y | WO 96 32504 A (UNIV BOSTON) 17 October 1996 see claims 1-27 | 1-44 |
| Y | WO 97 37041 A (SEQUENOM INC) 9 October 1997 see page 24, line 10 - line 27 | 1-44 |
| Y | EP 0 785 278 A (INST OF PHYSICAL & CHEMICAL RE) 23 July 1997 see the whole document | 1-44 |
| Y | US 5 571 669 A (LU PONZY ET AL) 5 November 1996 see page 1, column 1 - column 2, line 3 | 1-44 |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

I later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle of the underlying invention

- *X* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other cited documents, such combination being obvious to a person skilled in the art
- *S* document member of the same patent family

Date of the actual completion of the international search

23 April 1999

Date of mailing of the international search report

03/05/1999

Name and mailing address of the ISA
European Patent Office, P.O. Box 5818 Patentplan 2
NL - 8280 HV Tilburg
Tel. (+31-70) 343-2040, Tx. 31 651 eps nl.
Fax: (+31-70) 340-3016

Authorized officer

Osborne, H

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 98/26718

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| A | WO 89 07149 A (MAYO FOUNDATION) 10 August 1989 see the whole document ----- | 1-44 |
| A | WO 96 05323 A (CIS BIO INT ; SAUVAIGO SYLVIE (FR); BAZIN HERVE (FR); LIVACHE THIER) 22 February 1996 see abstract ----- | 1 |